

[illegible]

2

```

PCT-US01-16032-3
: Sequence 3, Application PC/UTS0116032
: GENERAL INFORMATION:
: APPLICANT: Cedars-Sinai Medical Center
: APPLICANT: The Regents of the University of California
: TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
: TITLE OF INVENTION: Crohn's Disease using The Ompc Antigen
: FILE REFERENCE: PP-PM 4713
: CURRENT APPLICATION NUMBER: PCT/US01/16032
: CURRENT FILING DATE: 2001-05-17
: PRIOR APPLICATION NUMBER: US 09/575,061
: PRIOR FILING DATE: 2000-05-19
: NUMBER OF SEQ ID NOS: 3
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 3
: LENGTH: 100
: TYPE: PRNT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Microbial organism from the human gut
PCT-US01-16032-3

```

Query Match	100.0%	Score 493;	DB 1;	length 100;
Best Local Similarity	100.0%	Pred. No. 4.6e-51;		
Matches 100; Conservative	0;	Mismatches	0;	Gaps 0

Oy 1 DLASAVGIOSSIFHHFKSKDEILRAVMEETIHYNTAMMRASLEASTYRERVLTALIRCE 600
Dd 1 DLASAVGIOSSIFHHFKSKDEILRAVMEETIHYNTAMMRASLEASTYRERVLTALIRCE 600

Qy	Db
61	61
10S1MGSGEAMAVLYEWMRSLSAEGGAHVLAIRDVEOI	10S1MGSGEAMAVLYEWMRSLSAEGGAHVLAIRDVEOI
1000	1000

3

```

PCT-US01-16032A-3
Sequence 3, Application PC/TUS0116032A
GENERAL INFORMATION:
APPLICANT: Targan, Stephan R.
APPLICANT: Braun, Jonathan
APPLICANT: Sutton, Christopher L.
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Crohn's Disease Using The OmpC Antigen
FILE REFERENCE: PP-PM 4713
CURRENT APPLICATION NUMBER: PCT/US01/16032A
CURRENT FILING DATE: 2001-05-17
PRIOR APPLICATION NUMBER: US 09/575,061
PRIOR FILING DATE: 2000-05-19
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 100
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Microbial organism from the human gut
PCT-US01-16032A-3

```

Query Match	100.0%	Score 493;	DB 1;	Length 100;
Best Local Similarity	100.0%	Pred. NO. 4.0e-51;		
Matches 100;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	DLASAVGIQGGSIFFHFRFSKDEILEAAVMEETHTHTNTAMKASLEEVSTYRKVALIRCE	60	

```
Db 1 DLSANGIQSGSTFHHFKKDELRLAWEHTIHTMTAMRSLEASTVERVALLICE 60
Qy 61 LQSMGSGSEAMAVLYEKRSLAEGQAHVLAIRVYEQI 100
    |||||
Db 61 LQSMGSGSEAMAVLYEKRSLAEGQAHVLAIRVYEQI 100
```

4

```

US-09-303-120B-2
: Sequence 2, Application US/09303120B
:
: GENERAL INFORMATION:
:
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher
: TITLE OF INVENTION: 1BD-Associated Microbial Antigens and Methods of Using
: TITLE OF INVENTION: Same
: FILE REFERENCE: P-PM 3478
: CURRENT APPLICATION NUMBER: US/09/303,120B
: CURRENT FILING DATE: 1999-04-30
:
: NUMBER OF SEQ ID NOS: 10
:
: SOFTWARE: PatentIn Ver. 2.1
:
: SEQ ID NO 2
:
: LENGTH: 100
:
: TYPE: PR1
:
: ORGANISM: Unknown
:
: FEATURE:
:
: OTHER INFORMATION: Description of Unknown Organism: Microbial
:
: US-09-303-120B-2

```

Query Match	100.0%	Score 493;	DB 17	Length 100;
Best Local Similarity	100.0%	Pred. No. 4.6e-51;		
Match Set 100; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

	61	LQSIMGSGEAMAVLYEWRSLSAEGCAHVLAIRDVEEOI	1000
Qy			
	61	LQSIMGSGEAMAVLYEWRSLSAEGCAHVLAIRDVEEOI	1000
Db			

5

```

RESULT      5
US-09-575-061-3
; Sequence 3, Application US/09573061
; GENERAL INFORMATION:
; APPLICANT: Targan, Stephan R.
; APPLICANT: Braun, Jonathan
; APPLICANT: Suton, Christopher L.
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; TITLE OF INVENTION: Crohn's Disease Using The Cmpc Antigen
; FILE REFERENCE: P-PW 4097
; CURRENT APPLICATION NUMBER: US/09/575,061
; CURRENT FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 100
; TYPE: PRF
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Microbial organism from the human gut
; US-09-575-061-3

```

```

Query Match      100.0% ; Score 493; DB 19; Length 100;
Best Local Similarity 100.0%; Pred. No. 4.ee-51;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLASAVGIOSCSIFHHKSKDEILRAVVEETHTNTAMRASLEASTVRRVALIRCE 60
   |||||
B 1 DLASAVGIOSCSIFHHKSKDEILRAVVEETHTNTAMRASLEASTVRRVALIRCE 60
```

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100
 DB 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

RESULT 6
 US-09-820-576-2

; Sequence 2, Application US/09820576
 ; GENERAL INFORMATION:

; APPLICANT: Braun, Jonathan

; TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid

; FILE REFERENCE: P-PM 4646

; CURRENT FILING DATE: 2001-03-28

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: FastSeq for Windows Version 4.0

; LENGTH: 100

; TYPE: PRT

; ORGANISM: Unknown

; OTHER INFORMATION: Microbial organism from the human gut

Query Match 100.0%; Score 493; DB 22; Length 100;
 Best Local Similarity 100.0%; Pred. No. 4.6e-51;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60
 DB 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100
 DB 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

RESULT 7
 US-09-252-991A-26292

; Sequence 26292, Application US/09252991A
 ; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT FILING DATE: 1999-02-18

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 26292

; LENGTH: 245

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

Query Match 87.0%; Score 429; DB 16; Length 245;
 Best Local Similarity 84.0%; Pred. No. 8.2e-43;
 Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60
 DB 87 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 146

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100

multiple 26292

me
 966
 500 10 #2

DNH
 24103

Sent for

DB 147 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 186

RESULT 8
 US-09-328-352-5572

; Sequence 5572, Application US/09328352
 ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

; FILE REFERENCE: GTC99-03PA

; CURRENT FILING DATE: 1999-06-04

; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 5572

; LENGTH: 206

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

Query Match 54.2%; Score 267; DB 17; Length 206;
 Best Local Similarity 54.0%; Pred. No. 2.2e-23;
 Matches 54; Conservative 21; Mismatches 25; Indels 0; Gaps 0;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 60
 DB 50 ELAOFGIOSGSLFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 109

QY 61 LOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 100
 DB 110 LISTGDTGAMAVLYEWRSLAEGQAVLALRDVEEI 149

RESULT 9
 PCT-US00-11473-6

; Sequence 6, Application PC/TUS0011473
 ; GENERAL INFORMATION:

; APPLICANT: The Regents of the University of California

; TITLE OF INVENTION: IBD-Associated Microbial Antigens and Methods of Using

; FILE REFERENCE: P-PM 4142

; CURRENT FILING DATE: 2000-04-28

; PRIOR FILING DATE: 1999-04-30

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patent In Ver. 2.0

; SEQ ID NO 6

; LENGTH: 190

; TYPE: PRT

; ORGANISM: Clostridium pasteurianum

Query Match 22.4%; Score 110.5; DB 1; Length 190;
 Best Local Similarity 27.0%; Pred. No. 0.00014;
 Matches 27; Conservative 25; Mismatches 35; Indels 13; Gaps 3;

QY 1 DIASAVGIOSGSIFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 56
 DB 28 EIASNAGVAKGTLFFHFKSKDELILAVMEETIHYNTAMRASLEASTYREVALIRCE 83

QY 57 INCELOSINGSGEAMAVLYEWRSLAEGQAVLALRDVEEI 96
 DB 84 CVOVNLILYKNDPFFKVIASQIMGR-----ELRQLELDI 118

RESULT 10
 US-09-303-120B-6
 ; Sequence 6, Application US/09303120B
 ; GENERAL INFORMATION:

```

RESULT 14
US-09-303-120B-7
Sequence 7, Application US/09303120B
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Sutton, Christopher
TITLE OF INVENTION: IBID-Associated Microbial Antigens and Methods of Using
FILE REFERENCE: P-PM 3478
CURRENT APPLICATION NUMBER: US/09/303,120B
CURRENT FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1

```


1. TITLE OF INVENTION: Crohn's Disease using Pseudomonas Antigens

104
114

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756

Query Match	98.28;	Score 484;	DB 5;	Length 198;
Best Local Similarity	98.08;	Pred. No. 8e-47;		
Matches	98;	Conservative	1;	Mismatches 0; Gaps 0;

QY	DB	QY	DB
1	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	61	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
2		62	
3		63	
4		64	
5		65	
6		66	
7		67	
8		68	
9		69	
10	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	70	
11		71	
12		72	
13		73	
14		74	
15		75	
16		76	
17		77	
18		78	
19		79	
20	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	80	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
21		81	
22		82	
23		83	
24		84	
25		85	
26		86	
27		87	
28		88	
29		89	
30	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	90	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
31		91	
32		92	
33		93	
34		94	
35		95	
36		96	
37		97	
38		98	
39		99	
40	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	100	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
41		101	
42		102	
43		103	
44		104	
45		105	
46		106	
47		107	
48		108	
49		109	
50	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	110	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
51		111	
52		112	
53		113	
54		114	
55		115	
56		116	
57		117	
58		118	
59		119	
60	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	120	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
61		121	
62		122	
63		123	
64		124	
65		125	
66		126	
67		127	
68		128	
69		129	
70	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	130	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
71		131	
72		132	
73		133	
74		134	
75		135	
76		136	
77		137	
78		138	
79		139	
80	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	140	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
81		141	
82		142	
83		143	
84		144	
85		145	
86		146	
87		147	
88		148	
89		149	
90	DLASVACIOGSGTFFHHFKSKDEILRAVMEETIHNVTAMRAMRSLEASTVREVALRICE	150	LASTMGSGEAMAVLYEWEKRSLSAEGQAHVALRDVYEOI
91			

```

RESULT      3
US-09-966-608-6
: Sequence 6, Application US/09966608
: GENERAL INFORMATION:
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher L.
: TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid
: TITLE OF INVENTION: Molecules
: FILE REFERENCE: P-PM 4966
: CURRENT APPLICATION NUMBER: US/09/966, 608
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: US 09/303,120
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: US 09/820,576
: PRIOR FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 190
: TYPE: PRF
: ORGANISM: Clostridium pasteurianum
US-09-966-608-6

```

	Query Match	Score	DB 5:	Length
Best Local Similarity	22.4%	110.5	DB 5:	190
Matches	27	Conservative	25	Mismatches 35; Indels 13; Gaps 3

```

RESULT      4
US-09-966-608-7
: Sequence 7, Application US/09966608
: GENERAL INFORMATION:
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher L.
: TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid
: TITLE OF INVENTION: Molecules
: FILE REFERENCE: P-PM 4966
: CURRENT APPLICATION NUMBER: US/09/966, 608

```

```

:
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: US 09/303,120
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: US 09/820,576
: PRIOR FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 7
: LENGTH: 200
: TYPE: PAT
: ORGANISM: Mycobacterium tuberculosis
: US-09-966-608-7

```

Query Match	19.28;	Score	94.5;	DB	5;	Length	200;
Best Local Similarity	26.98;	Pred. No.	0.0045;				
Matches	21;	Conservative	20;	Mismatches	34;	Indels	3;
						Gaps	1;

```

QY 1 DLSAVGIOGSGSFFHFKRCKDELLAVMEETJYNTNPMRARSLEASTVERVIALRCE 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 35 DIADGACILGSLYHNFASKKEEWDELLRGFLMDLFARYRDIVSTANPLERLOGGFMAS 90
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 LQSGINGSGEAMAVLYE 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 95 FEATEHHNAO--VVIQ 109
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT      5
US-09-897-516-4731
Sequence 4731, Application US/09897516
GENERAL INFORMATION:
APPLICANT: Corbin, David R.
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Huesing, Joseph E.
APPLICANT: Krasomil-Osterfeld, Karina C.
APPLICANT: Malvar, Thomas M.
APPLICANT: Slater, Steven C.
APPLICANT: Spirdonov, Sergei
TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
FILE REFERENCE: 38-21(51847)B
CURRENT APPLICATION NUMBER: US/09/897,516
CURRENT FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215, 161
NUMBER OF SEQ ID NOS: 8409
SEQ ID NO 4731
LENGTH: 193
TYPE: PRF
ORGANISM: Xenorhabdus sp.
US-09-897-516-4731

```

[illegible]

```
RESULT 6
US-09-605-703B-2830
: Sequence 2830, Application US/09605703B
: GENERAL INFORMATION:
: APPLICANT: Pompejus, Markus
: APPLICANT: Krogger, Burkhard
: APPLICANT: Schroder, Hartwig
```



```

1  APPLICANT: Zelder, Oskar
2  APPLICANT: Haberhauser, Gregor
3  TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING NOVEL
4  TITLE OF INVENTION: PROTEINS
5  FILE REFERENCE: BGI-129CP
6  CURRENT APPLICATION NUMBER: US/09/605,703B
7  CURRENT FILING DATE: 2000-06-27
8  PRIOR APPLICATION NUMBER: 60/142,764
9  PRIOR FILING DATE: 1999-07-08
10 PRIOR APPLICATION NUMBER: 60/152,318
11 PRIOR FILING DATE: 1999-09-03
12 NUMBER OF SEQ ID NOS: 2934
13 SEQ ID NO 2830
14 LENGTH: 218
15 TYPE: PRP
16 ORGANISM: Corynebacterium glutamicum
17 US-09-605-703B-2830

```

Query Match	18.3%	Score 90;	DB 5;	Length 218;
Best Local Similarity	29.4%	Pred. No. 0.016;		
Matches	20;	Conservative	20;	Indels 8; Gaps 2

```

Oy      2 LASANGIQSGSLFHFKSKDEILRAVMEETHYNTAMRASLEASTV----RERYLALJ 57
      :| ||||| :||| ||| :||| :||| :||| :||| :||| :|||
Db      41 IADAVGIRQASLYYHFPSPKTEIFLTLLKSTYEPSTVL---AEDSLDAGPEMRMAIV 96

```

QY	58	RCELOSIM	65
	1:::	:	:
Db	97	ASEVRLL	104

```

7
US-09-966-608-8
: Sequence 8, Application US/09966608
: GENERAL INFORMATION:
: APPLICANT: Braun, Jonathan
: APPLICANT: Sutton, Christopher L.
: TITLE OF INVENTION: IBD-Associated Microbial Nucleic Acid
: TITLE OF INVENTION: Molecules
: FILE REFERENCE: P-PM 4966
: CURRENT APPLICATION NUMBER: US/09/966, 608
: CURRENT FILING DATE: 2001-09-27
: PRIOR APPLICATION NUMBER: US 09/303,120
: PRIOR FILING DATE: 1999-04-30
: PRIOR APPLICATION NUMBER: 09/820,576
: PRIOR FILING DATE: 2001-03-28
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8
: LENGTH: 192
: TYPE: PRT
: ORGANISM: Aulifex aeolicus
US-09-966-608-8

```

Query Match	18.1%	Score	89	DB	5	Length	192
Best Local Similarity	41.3%	Pred	No.	0.017			
Matches	19	Conservative	10	Mismatches	13	Indels	4
						Gaps	1

```
Oy      1 DLASAVGIQSGSIFHHFKSKDEILRAVMEETIHYNTAMRASLEEA 46
        | : | | | : | | : | : : | | : | | |
Db     38 DIAKEVGITEGAIYRHFTSKEEIIKSLLESI ---TKELRHKLEVA 79
```

RESULT 8
US-08-895-611-2
Sequence 2, Application US/08895611
GENERAL INFORMATION:
APPLICANT: DIVERSA CORPORATION
APPLICANT: Swanson, Ronald V.
TITLE OF INVENTION: Short, Jay M.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF THE AQUIFER

```

: TITLE OF INVENTION: ABOULICUS GENOME, FRAGMENTS THEREOF, AND USES THEREOF
:
: FILE REFERENCE: DIYER1330
:
: CURRENT APPLICATION NUMBER: US/08/895,611
:
: CURRENT FILING DATE: 1997-07-16
:
: NUMBER OF SEQ ID NOS: 5
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO: 2
:
: LENGTH: 576775
:
: TYPE: PR1
:
: ORGANISM: Aquifex aeolicus
:
: US-08-895-611-2

```

Query Match	18.1%	Score 89:	DB 4:	Length 576775:
Best Local Similarity	41.3%	Pred. No.	7.6e+02:	
Matches 19:	Conservative 10:	Mismatches 13:	Indels 4:	Gaps 1:

```
QY      1 DLASAVGIQSGSIFHHFKSKDEILRAVMEETHYNTAMMRASLEEA 46
      1:| 1| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:| 1:|
Db 365113 DIAKEVGITEGAIYRHFPTSKEEIIKSLLESI---TKELRHKLEVA 365154
```

```

RESULT      9
US-08-895-611D-2
: Sequence 2, Application US/08895611D
: GENERAL INFORMATION:
: APPLICANT: DIVERSA CORPORATION
: APPLICANT: Swanson, Ronald V.
: APPLICANT: Short, Jay M.
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF THE AQUIFEX
: FILE REFERENCE: DIVERS130
: CURRENT APPLICATION NUMBER: US/08/895,611D
: CURRENT FILING DATE: 1997-07-16
: NUMBER OF SEQ ID NOS: 5
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 576775
: TYPE: PRF
: ORGANISM: Aquifex aeolicus
US-08-895-611D-2

```

Query Match	18.1%	Score	89	DB	4	Length	576775
Best Local Similarity	41.3%	Pred. No.	7.6e+02				
Matches	19	Conservative	10	Mismatches	13	Indels	4
						Gaps	1

```
Oy      1 DLASAVGIOSGSIFHHFKSKDEILRAVMEETIHYNTAMRASLEEA 46
        |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:| |:|
Db 365113 DIAKEVGITEGAIYRHFTSKEEIIKSLEST-----TKELRHKLEVA 365154
```

```

RESULT 10
US-09-895-611D-2
; Sequence 2, Application US/09895611D
; GENERAL INFORMATION:
; APPLICANT: DIVERSA CORPORATION
; APPLICANT: Swanson, Ronald V.
; APPLICANT: Short, Jay M.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF THE AQUIFEX
; TITLE OF INVENTION: AQUIFEX GENOME, FRAGMENTS THEREOF, AND USES THEREOF
; FILE REFERENCE: DIVER1330
; CURRENT APPLICATION NUMBER: US/09/895.611D
; CURRENT FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 576775
; TYPE: PR1
; ORGANISM: Aquifex aeolicus
US-09-895-611D-2

```

Query Match	17.0%	Score 84	DB 5	Length 196																																												
Best Local Similarity	29.8%	Pred. No. 0.065																																														
Matches 17; Conservative 16; Mismatches 22; Indels 2; Gaps 1;																																																
QY	1	D	L	A	S	A	V	G	I	Q	S	S	T	R	H	E	R	K	S	K	D	E	L	A	V	E	E	T	I	H	T	A	M	R	S	L	E	A	S	T	R	--	E	R	V	L	A	55
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:			
	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:				
35	E	I	C	A	A	M	S	P	E	N	E	L	H	Y	P	T	K	N	A	I	L	E	A	E	E	S	H	D	E	L	I	A	C	D	E	N	S	T	M	T	E	K	L	A	91			

```

Query March 16.1% Score 79.5: DB 5. Length 193:
Best Local Similarity 27.0% Pred. No. 0.21:
Matches 27: Conservative 18: Mismatches 50: Indels 5: Gaps 3:

2 LASAVGDSGSIPIHFHKSDKELRAVMEETH--YNTAMRASLEASTVERRYALALRC 59
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
35 IAKKAGVSGNLSIHFKKRNKLLLEAMRHLLHQLOMAVARRLRLDNDTPRLRLAIEG 94

```

OY 60 ELOSINGSGEAMAVLYEWMRELSAEGOAHVLAIRDVEO 99
DB 95 NFDTSQINSA-AMKTWIAFW--ASSMHQPSLYRLQOVNER 131

RESULT 15
US-09-815-242-5102

; Sequence 5102, Application US/09815242
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5102
; LENGTH: 194
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5102

Query Match 15.7%; Score 77.5; DB 5; Length 194;
Best Local Similarity 25.0%; Pred. No. 0.35;
Matches 25; Conservative 23; Mismatches 43; Indels 9; Gaps 3;

OY 1 DIASAVGIQSGSIFHFPSKDELIRAVMEETH-INTAMRASLEASTVREPVIALIRC 59
DB 32 EILQSGAGVPGSGFYHFFKSKQEGQALLDYPFVYLAIDMDQRFSAAGLNARERLMSYQK 91
OY 60 ELOSINGSGEAMAVLYEWMRELSAE---GOAHVLAIRD 95
DB 92 WIDNACPCDEORCLV---KLSAEVADLSESMRITLRD 127

Search completed: March 4, 2002, 20:31:16
Job time: 399 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:22:32 ; Search time 27.71 Seconds
(without alignments)
267.316 Million cell updates/sec

Title: US-09-966-608-2
Perfect score: 493
Sequence: 1 DLASAVGIGSGSIFHFKSK.....SLSAEGCAHYLALRDVEQL 100

Scoring table: BLOSUM62
Gap 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	493	100.0	100	21	AA19839
2	90	18.3	222	22	AA19839
3	83	16.8	214	22	AA19839
4	78.5	15.9	118	21	AA19839
5	73	14.8	3722	12	AA19839
6	70.5	14.3	188	22	AA19839
7	69	14.0	80	21	AA19839
8	69	14.0	2818	13	AA19839
9	68	13.8	509	19	AA19839
10	68	13.8	509	19	AA19839
11	68	13.8	509	21	AA19839

12	67.5	13.7	508	17	AA19839	Protein having pro
13	67.5	13.7	508	18	AA19839	Arabidopsis thalia
14	67.5	13.7	508	18	AA19839	Arabidopsis thalia
15	67.5	13.7	508	19	AA19839	Arabidopsis thalia
16	67.5	13.7	508	19	AA19839	Arabidopsis thalia
17	67.5	13.7	508	22	AA19839	Arabidopsis thalia
18	66.5	13.5	136	22	AA19839	Arabidopsis thalia
19	66.5	13.5	195	19	AA19839	Arabidopsis thalia
20	65.5	13.3	411	19	AA19839	Arabidopsis thalia
21	65.5	13.3	414	19	AA19839	Arabidopsis thalia
22	65.5	13.3	444	19	AA19839	Arabidopsis thalia
23	64.5	13.1	195	21	AA19839	Arabidopsis thalia
24	64.5	13.1	326	21	AA19839	Arabidopsis thalia
25	64	13.0	156	21	AA19839	Arabidopsis thalia
26	64	13.0	783	22	AA19839	Arabidopsis thalia
27	64	13.0	3768	12	AA19839	Arabidopsis thalia
28	64	13.0	3778	12	AA19839	Arabidopsis thalia
29	63.5	12.9	1784	16	AA19839	Arabidopsis thalia
30	63	12.8	2485	15	AA19839	Arabidopsis thalia
31	63	12.8	2485	15	AA19839	Arabidopsis thalia
32	63	12.8	2818	18	AA19839	Arabidopsis thalia
33	63	12.8	2818	22	AA19839	Arabidopsis thalia
34	62.5	12.7	82	21	AA19839	Arabidopsis thalia
35	62.5	12.7	224	21	AA19839	Arabidopsis thalia
36	62	12.6	76	22	AA19839	Arabidopsis thalia
37	62	12.6	217	22	AA19839	Arabidopsis thalia
38	62	12.6	219	22	AA19839	Arabidopsis thalia
39	62	12.6	435	22	AA19839	Arabidopsis thalia
40	62	12.6	483	22	AA19839	Arabidopsis thalia
41	62	12.6	778	21	AA19839	Arabidopsis thalia
42	62	12.6	806	21	AA19839	Arabidopsis thalia
43	62	12.6	927	21	AA19839	Arabidopsis thalia
44	61.5	12.5	76	14	AA19839	Arabidopsis thalia
45	61.5	12.5	211	22	AA19839	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AA19839	standard; Protein: 100 AA.
ID	AA19839	
AC	AA19839	
XX		
DT	05-MAR-2001	(first entry)
XX		
DE	Inflammatory bowel disease associated antigen I-2.	
XX		
KW	I-2; Inflammatory bowel disease; IBD; Crohn's disease; therapy; diagnosis; vaccine.	
XX		
OS	Unidentified microorganism.	
XX		
PN	WO200066067-A2.	
XX		
PD	09-NOV-2000.	
XX		
PF	28-APR-2000; 2000MO-US11473.	
XX		
PR	30-APR-1999; 99US-0303120.	
XX		
PA	(RECC) UNTV CALIFORNIA.	
XX		
PI	Braun J, Sutton C;	
XX		
DR	WPI: 2000-687440/67.	
XX		
DR	N-PSDB: AAA89029.	
XX		
PT	Inflammatory bowel disease (IBD) associated I-2 polypeptides useful for	
PT	diagnosing IBD, vaccinating against IBD and for identifying agents for	
PT	treating IBD -	
XX		

PS Claim 1; Fig 1A; 76pp; English.

XX The present sequence is that of novel microbial I-2 polypeptide.

CC Representative difference analysis was used to isolate DNA

CC sequences from a Crohn's disease (CD) patient that were

CC differentially present in mononuclear cells from the lamina propria

CC in an area with ulcerations as compared to an area free of disease.

CC 2 inflammatory bowel disease (IBD)-associated sequences were of

CC microbial origin, designated I-1 (see AAB19840-41) and I-2. I-2

CC sequences were more often found in involved CD tissue than in

CC ulcerative colitis or non-IBD samples. The I-2 polypeptide may be

CC used in methods of the invention for treating IBD, diagnosing IBD,

CC vaccinating against IBD, and identifying candidate agents useful

CC for treating IBD, especially CD.

XX

SQ Sequence 100 AA;

Query Match 100.0%; Score 493; DB 21; Length 100;

Best Local Similarity 100.0%; Pred. No. 1.4e-54;

Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DLASAVGIQSGSIFPHFKSKDEILRAVMEETIHYNTAMMRASLEASVREVALIRCE 60

Db 1 dlasevgtqsgsifhfkskdeilravmeethyntamtrasleasvrevallirce 60

QY 61 LOSINGSGEAMAVLYEMRSLAEGQAVLALRDVYEOI 100

Db 61 lqslngsggeamavlyewrslseagqavhlalrdvyeqi 100

RESULT 2

AA090710 standard; Protein: 222 AA.

XX

AC AAG90710;

XX

DT 26-SEP-2001 (first entry)

XX

DE C glutamylum protein fragment SEQ ID NO: 4464.

XX

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

XX

KM organic acid synthesis.

XX

OS Corynebacterium glutamicum.

XX

PN EPI108790-A2.

XX

PD 20-JUN-2001.

XX

PF 18-DEC-2000; 2000EP-0127688.

XX

PR 16-DEC-1999; 99JP-0377484.

XX

PR 07-APR-2000; 2000JP-0159162.

XX

PR 03-AUG-2000; 2000JP-0280988.

XX

PA (KYOW) KYOWA HAKKO KOGYO KK.

XX

PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX

PI Tereishi N, Senoh A, Ikeda M, Ozaki A;

XX

DR WPI: 2001-376931/40.

XX

DR N-PSDB: AAH65929.

XX

PT Novel polynucleotides derived from Coryneform bacteria, for identifying

PT mutation point of a gene, measuring expression of a gene, analysing

PT expression profile or pattern of a gene and identifying homologous gene

XX

PS Claim 17; SEQ ID NO: 4464; 246pp + sequence listing; English.

XX

CC The present invention provides a number of nucleotide and protein

CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX

SQ Sequence 222 AA;

Query Match 18.3%; Score 90; DB 22; Length 222;

Best Local Similarity 29.4%; Pred. No. 0.0028;

Matches 20; Conservative 20; Mismatches 20; Indels 8; Gaps 2;

QY 2 LASAVGIQSGSIFPHFKSKDEILRAVMEETIHYNTAMMRASLEASVREVALIR 57

Db 45 ladevgtqsgsifhfkskdeilravmeethyntamtrasleasvrevallirce 100

QY 58 RCEIQTSM 65

Db 101 asevrlll 108

RESULT 3

AA004043 standard; Protein: 214 AA.

XX

AC AAU04043;

XX

DT 23-OCV-2001 (first entry)

XX

DE Streptococcus coelicolor Mmfr protein.

XX

KW SCP1; methylenomycin cluster; mmc; MmyR; Mmfp; Mmfi;

XX

KM Mmfl; Mmfr; MmyT; Mmyo; MmyG; MmyJ; Mmr; heterologous gene expression.

XX

OS Streptococcus coelicolor.

XX

PN WO200148228-A1.

XX

PD 05-JUL-2001.

XX

PF 20-DEC-2000; 2000MO-GB04972.

XX

PR 23-DEC-1999; 99GB-0030477.

XX

PA (PLAN-) PLANT BIOSCIENCE LTD.

XX

PI Chater KF, Bruton CJ, O'Rourke SJ, Wietzorrek AW;

XX

XX WPI: 2001-425675/45.

XX

DR N-PSDB: AAS07627.

XX

DR AAU04045, AAU04046.

XX

PT Novel expression cassette for expressing a nucleic acid of interest,

PT derived from the regulatory region of methylenomycin gene cluster of

PT SCP1 plasmid of Streptomyces coelicolor A3(2)

XX

PS Claim 25; Fig 8e; 142pp; English.

XX

CC The sequence represents the Mmfr protein encoded by the mmfr gene carried

CC on the expression cassette present on plasmid SCP1. The expression

CC cassette is the regulatory region of the methylenomycin cluster (mmc)

CC from Streptomyces coelicolor A3(2), which encodes the MmyR, Mmfp, Mmfi,

CC Mmfl, Mmfr, MmyT, Mmyo, MmyG, MmyJ and partial Mmr polypeptides. The

CC expression cassette is useful for expressing a nucleic acid of interest,

CC substantially only when the host cell culture reaches high cell density

CC at or close to the stationary phase of host cell culture. In particular

AAC/44446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,

PT prepn. of cephalosporin series antibiotics - comprises culturing
PT transformant of microbe transformed by plasmid contg. new DNA
PT fragment

PS Disclosure: Fig 13; 67pp; Japanese.
 CC This protein is encoded by ORF1 of the 23666bp sequence
 CC isolated from *L. lactamgenus* and comprising the genes for the
 CC cephalosporin biosynthetic enzymes listed in the KEYWORDS. Plasmids
 CC containing at least one of ORF's 1-9 can be used to transform
 CC microbes, such as bacteria or yeast.
 CC See also AAO10191-2.
 CC
 XX Sequence 3722 AA;
 SQ
 Query Match 14.8%; Score 73; DB 12; Length 3722;
 Best Local Similarity 26.3%; Pred. NO. 18;
 Matches 20; Conservative 12; Mismatches 28; Indels 16; Gaps 1;
 OY 5 AVGTSGSIFPHFKSDEILRAVMEETIHYNTAMRASLEASTVRRVRLALIRCELOSI 64
 DB 901 atg1qggmlygsmkssandayvmqslhryraldpaameqawlaqgkypslr----- 954
 OY 65 MCGSGEAMAVLYEWR 80
 DB 955 -----lrfewr 960
 RESULT 6
 AAG91446
 ID AAG91446 standard; Protein; 188 AA.
 XX
 XX AAG91446;
 XX
 XX 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 5200.
 XX
 KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 XX EPI108790-A2.
 XX
 XX 20-JUN-2001.
 XX
 XX 18-DEC-2000; 2000EP-0127688.
 XX
 XX 16-DEC-1999; 99JP-0377484.
 XX
 XX 07-APR-2000; 2000JP-0159162.
 XX
 XX 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOM) KYOMA HAKKO KOGYO KK.
 XX
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 XX Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 XX
 XX N-PSDB: AAH66665.
 XX
 DR Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 XX
 PS Claim 17; SEQ ID NO: 5200; 246pp + Sequence Listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium *Corynebacterium glutamicum*. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX Sequence 188 AA;
 SQ
 Query Match 14.3%; Score 70.5; DB 22; Length 188;
 Best Local Similarity 25.8%; Pred. NO. 0.66;
 Matches 24; Conservative 9; Mismatches 23; Indels 37; Gaps 4;
 OY 2 LASAVGTSGSIFPHFKSDEILRAV-----MEETIHYN--TAMRASLEASTVRR 52
 DB 37 leeatgksrgaifhbgdkenlfalaredaarmaevsenglvemrgmledp-----er 92
 OY 53 VLALIRCELOISMGSGEAMAVLYEWRSLSAE 85
 DB 93 -----ydwmsvrle 101
 RESULT 7
 AAB44577
 ID AAB44577 standard; Protein; 80 AA.
 XX
 XX AAB44577;
 XX
 XX 08-FEB-2001 (first entry)
 XX
 DE Virulence gene protein #57.
 XX
 XX Virulence gene; antibacterial; vaccine; bacterial infection;
 XX septicemia; bronchopneumonia; rhinitis; wound infection.
 XX
 KM Actinobacillus pleuropneumoniae.
 XX
 OS WO200061724-A2.
 XX
 XX 19-OCT-2000.
 XX
 XX 06-APR-2000; 2000WO-US09218.
 XX
 XX 09-APR-1999; 99US-0128689.
 XX
 XX 10-SEP-1999; 99US-0153453.
 XX
 XX (PHAA) PHARMACIA & UPJOHN INC.
 XX
 XX Lowery DE, Fuller TE, Kennedy MJ;
 XX
 XX WPI: 2000-647422/62.
 XX
 XX N-PSDB: AAC79652.
 XX
 DR Attenuated Pasteurellaceae bacteria comprising mutations in virulence
 XX genes, useful as a live attenuated vaccine against bacterial infections
 XX
 XX Claim 39; Page 288; 322pp; English.
 XX
 XX The family Pasteurellaceae encompasses several pathogens that infect a
 CC wide variety of animals. The present invention relates to virulence genes
 CC from Pasteurellaceae. The present sequence is a protein encoded by one
 CC such virulence gene. The virulence genes of the present invention may be
 CC mutated in order to produce an inactive gene. The inactive virulence gene
 CC may in turn be used to produce a vaccine, which is useful for treating
 CC bacterial infections such as septicemias, bronchopneumonias, rhinitis and
 CC wound infections.
 XX
 SQ Sequence 80 AA;
 Query Match 14.0%; Score 69; DB 21; Length 80;
 Best Local Similarity 41.4%; Pred. NO. 0.32;

[illegible]

Pf		10-FEB-2000;	2000OWO-US03381.	xx
Xx				
Pt		02-MAR-1999;	99US-0122499.	xx
Xx				
Pa	(REGE -)	REGENERON PHARM INC.		xx
Xx	Jones PF,	Valenzuela DM;		xx
Pi				
Dn	WPI:	2000-579286/54.		xx
Dr	N-PSTDB:	AHAJ37838.		
Xx	Novel nucleic acid molecules encoding fibrinogen-like domain of Angiopietein Related-2 factor useful for treating Ischaemia, diabetes, and for wound healing	-		xx
Pt				
Xx				
Pt				
Xx	Example 1; Fig 1; 55pp;	English.		xx
Xx	This sequence is the mouse Tie ligand-3. This sequence was used in the isolation of the angiotensin related-2 (Ar-2) fibrinogen-like domain (FP) of the invention. Ar-2 is useful for treating and diagnosing ischaemia, diabetes, tumour angiogenesis, neoplastic diseases, thromboembolic diseases, arteriosclerosis, inflammatory diseases, and for wound healing. The DNA is useful for developing ligands, screening agonists and antagonists of Ar-2, and as a therapeutic for treating disorders involving cells, tissues or organs expressing Ar-2 receptor. Ar-2 is useful to promote the growth, survival, migration, stabilisation or destabilisation, and/or differentiation of cells expressing Ar-2 receptor. Ar-2 is also useful in assay systems to identify agonists and antagonists of Ar-2 receptor. Ar-2 is also useful for inducing or preventing vascularisation in diseases or disorders where such function is indicated, for delivering toxins to a receptor bearing cells and as diagnostic reagents for detecting the disease by tissue staining or whole body imaging.			xx
Sd	Sequence	509 AA;		
Qy	Query Match	13.8%; Score 66; DB 21; Length 509;		
Bst	Best Local Similarity	28.4%; Pred. NO. 5.2;		
Mch	Matches	21; Conservative 18; Mismatches 29; Indels 6; Gaps 2;		
Db	165 VINQLIMTKTQMISLEASTVRRERVALIRCELOSIMGSGGEMAVLYEWMRSLSAEC	86		
Qy	87 QAHVLALRDYVEOI	100		
Db	219 qadlnslqekreql	232		
Rslt	RESULT 12			
ID	AAR90296	standard; protein; 508 AA.		
Cc	AAR90296;			
Dt	02-JUL-1996	(first entry)		
Xx	Protein having protoporpyrinogen oxidase activity.			
Xx	Protoporphyrinogen oxidase; PRO; herbicide; resistance; mutant variegate porphyrila.			
Xx	Arabidopsis sp.			
OS	WO9534659-A1.			
PN	21-DEC-1995.			
PD	08-JUN-1995;	95WO-TB00452.		
PF	16-JUN-1994;	94US-0261198.		
PR				

	PA	(CIBA) CIBA GEIGY AG.				
Pt	PI	Volrath S., Ward ER;				
Df	DR	WPI; 1996-049687/05.				
Xx	DR	N-PSDB; AAT11674.				
Pt	XX	DNA encoding eukaryotic proto.porphyrinogen oxidase and herbicide resistant mutants - used to make herbicide resistant plants and for diagnosis and treatment of variegate porphyria				
Pt	PR	Claim 6; Page 77-79; 118pp: English.				
Pt	PP	DNA encoding a modified protoporphyrinogen oxidase (PPO) or a chimeric gene comprising a promoter (pref. active in a plant) linked to such a DNA or DNA encoding a wild type PPO can be used (1) to impart herbicide resistance to plants; (2) for treating and diagnosing deficient PPO activity in animals (esp. variegated porphyrina); and (3) for the production of recombinant PPO which is useful as an assay reagent and in rational design of new inhibitory herbicides. Herbicide resistant PPO genes can also be used to select plants transformed with a transgene and probes derived from the genes can be used to quantify levels of PPO mRNA.				
SQ	XX	Sequence 508 AA:				
	Query Match	Best Local Similarity 13.7%; Score 67.5; DB 17; Length 508; Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3				
Oy	15	HHEFKSKDELRAVM---EEFIHYNTAMRRAS---LEEAFTVERVUALLRCLOSLIM 65 : : : : : :: :::: :: :				
Db	366	hgkftlgltffssmfpdrpsdvhytltfifggsrngelakast--dclqyvtsldlgrll 423 : : : : : :				
Oy	66	GGSCEAAMVLVEWR 80 : : :				
Db	424	fvegepvsvnyhywr 438				
	RESULT 13					
ID	AAM41604	standard; protein: 508 AA.				
AC	AAM41604;					
DF	20-APR-1998	(first entry)				
DE	Arabidopsis thaliana ptrox-2.					
KM	Protoporphyrinogen oxidase-2; ptrox-2; promoter; herbicide resistance; breeding programme; probe; gene isolation; genomic mapping.					
OS	Arabidopsis thaliana.					
PN	WO9732028-A1.					
PD	04-SEP-1997.					
PF	27-FEB-1997; 97WO-USO3343.					
PR	21-JUN-1996; 96US-0020003. 28-FEB-1996; 96US-0012705. 28-FEB-1996; 96US-0013612.					
PA	(NOVS) NOVARTIS AG.					
PI	Johnson MA., Volrath SL., Ward ER;					
DR	WPI; 1997-489209/45.					
DR	N-PSDB; AAV04305.					
XX						

```

PT DNA containing a plant proto-porphyrinogen oxidase gene promoter
PT optionally linked to a heterologous gene, especially to express
PT herbicide-resistant enzymes, and plants containing such constructs
XX
PS Claim 27; Pages 43-46; 114pp; English.
XX
CC The present sequence is Arabidopsis thaliana
CC protoporphyrinogen oxidase-2 (protox-2).
CC The protox-1 promoter can be used to express herbicide resistant
CC enzymes, specifically protox, i.e. a plant tissue, plant or progeny
CC containing a chimeric gene of the promoter and a heterologous
CC coding sequence. The plant can also be used in breeding programmes.
CC Also hybridising fragments of the protox coding sequence can be
CC used as probes, e.g. to isolate related genes or for genomic
CC mapping.
CC
XX Sequence 508 AA;
XX
SQ
Query Match 13.7%; Score 67.5; DB 18; Length 508;
Best Local Similarity 26.7%; Pred. No. 6;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3
OY 15 HHKSKDELRAVM-----EETIHYNTAMRAS---LEASTVRRRLALIRCELOSIM 65
Db 366 hqikcltqltffssmfpdrpsdvhlytftlfgsrrngelakast--delkqvslsdqrl 423
OY 66 GSGSEAMAVLVYEMR 80
Db 424 gvegepvsynhywr 438
RESULT 14
AAM25747
ID AAM25747 standard; Protein: 508 AA.
XX AAM25747;
AC
XX
XX 01-MAR-1998 (first entry)
DT
XX
XX Arabidopsis protoporphyrinogen oxidase (protox-2).
DE
XX
XX Arabidopsis protoporphyrinogen oxidase; inhibitor;
KM herbicide tolerance; herbicide resistance; transgenic plant.
XX
XX Arabidopsis thaliana.
OS
XX
XX WO9732011-A1.
PN
XX
XX 04-SEP-1997.
PD
XX
XX 27-FEB-1997; 97WO-US03313.
PF
XX
XX 21-JUN-1996; 96US-0020003.
PR 28-FEB-1996; 96US-0012705.
PR 28-FEB-1996; 96US-0013612.
XX
XX (NOVS ) NOVARTIS AG.
PA
XX
XX Helfetz PR, Johnson MA, Potter SL, Volrath SL, Ward ER;
XX
XX WPI: 1997-448683/41.
DR N-PSDB: AAT86130.
XX
XX New DNA encoding plant protoporphyrinogen oxidase enzyme - and
PT herbicide resistant mutants; useful to prepare plants resistant to
PT herbicide which therefore kills undesired vegetation only
XX
XX Claim 95; Page 102-105; 196pp; English.
XX
CC This protein comprises Arabidopsis protoporphyrinogen oxidase
CC (protox-2), an enzyme that catalyses the oxidation of
CC protoporphyrinogen IX to protoporphyrin IX. Its amino acid sequence

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:23:12 ; Search time 15.26 Seconds
(without alignments)
147.466 Million cell updates/sec

Title: US-09-966-608-2
Perfect score: 493
Sequence: 1 DLASAVGIGSGSIFHHFKSK.....SLSAEQGAHVLRDVEQI 100

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	13.8	503	4	US-08-740-223A-11 Sequence 11, Appl
2	68	13.8	509	2	US-08-665-926-8 Sequence 8, Appl
3	68	13.8	509	4	US-08-740-223A-10 Sequence 10, Appl
4	67.5	13.7	508	1	US-08-472-028A-4 Sequence 4, Appl
5	67.5	13.7	508	2	US-08-808-931-4 Sequence 4, Appl
6	67.5	13.7	508	3	US-08-808-323-4 Sequence 4, Appl
7	67.5	13.7	508	3	US-09-050-603A-4 Sequence 4, Appl
8	67.5	13.7	508	3	US-09-102-420B-4 Sequence 4, Appl
9	67.5	13.7	508	4	US-09-071-296-4 Sequence 4, Appl
10	67.5	13.7	508	4	US-09-196-268-4 Sequence 4, Appl
11	67.5	13.7	508	4	US-09-015-683-4 Sequence 4, Appl
12	67.5	13.7	508	4	US-09-191-998-4 Sequence 4, Appl
13	67	13.6	2818	2	US-08-449-933-2 Sequence 2, Appl
14	67	13.6	2818	4	US-07-966-049A-2 Sequence 2, Appl
15	65.5	13.3	195	2	US-08-882-704A-2 Sequence 2, Appl
16	64	13.0	768	2	US-08-222-617A-5 Sequence 5, Appl
17	64	13.0	3666	2	US-08-222-617A-12 Sequence 12, Appl
18	64	13.0	3727	2	US-08-222-617A-27 Sequence 27, Appl
19	64	13.0	3778	2	US-08-222-617A-2 Sequence 2, Appl
20	63.5	12.9	1784	4	US-09-040-738-2 Sequence 2, Appl
21	63.5	12.9	1784	4	US-08-652-426A-2 Sequence 2, Appl
22	63	12.8	2485	5	PCT-US94-00198-1 Sequence 1, Appl
23	63	12.8	2485	5	PCT-US94-00198-2 Sequence 2, Appl
24	63	12.8	2818	1	US-08-510-284-1 Sequence 1, Appl
25	63	12.8	2818	1	US-08-411-389-2 Sequence 2, Appl
26	63	12.8	2818	1	US-09-542-331-2 Sequence 2, Appl
27	62	12.6	219	4	US-09-162-184-32 Sequence 32, Appl

28	62	12.6	219	4	US-09-489-777A-32 Sequence 32, Appl
29	61.5	12.5	76	2	US-08-244-537-6 Sequence 6, Appl
30	61.5	12.5	2325	3	US-08-417-089-6 Sequence 6, Appl
31	61.5	12.5	2325	4	US-08-695-651-6 Sequence 6, Appl
32	61.5	12.5	2325	4	US-08-930-285-6 Sequence 6, Appl
33	61.5	12.5	2325	4	US-08-695-421-6 Sequence 6, Appl
34	59.5	12.1	216	4	US-09-162-184-31 Sequence 31, Appl
35	59.5	12.1	216	4	US-09-489-777A-31 Sequence 31, Appl
36	59	12.0	218	4	US-09-162-184-30 Sequence 30, Appl
37	59	12.0	218	4	US-09-489-777A-30 Sequence 30, Appl
38	58.5	11.9	785	1	US-07-841-997A-4 Sequence 4, Appl
39	58.5	11.9	785	1	US-08-290-301-4 Sequence 4, Appl
40	58.5	11.9	1098	1	US-08-290-301-82 Sequence 82, Appl
41	58	11.8	295	1	US-08-317-223-1 Sequence 1, Appl
42	58	11.8	295	3	US-09-059-849A-1 Sequence 1, Appl
43	58	11.8	295	3	US-09-213-632-1 Sequence 1, Appl
44	58	11.8	295	5	PCT-US95-12675-1 Sequence 1, Appl
45	58	11.8	770	1	US-08-445-135-2 Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-740-223A-11
Sequence 11, Application US/08740223A
Patent No. 6265564
GENERAL INFORMATION:
APPLICANT: Davis, et al.
TITLE OF INVENTION: Expressed Ligand - Vascular
TITLE OF INVENTION: Intercellular Signalling Molecule
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneon Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill Road
CITY: Tarrytown
STATE: NY
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/740,223A
FILING DATE: 25-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/022/999
FILING DATE: 02-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Coberl, Robert J
REGISTRATION NUMBER: 36,108
REFERENCE/DOCKET NUMBER: REG 333
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: mtl3
LOCATION: 1...503
OTHER INFORMATION: mouse TIE ligand-3
US-08-740-223A-11
Query Match 13.8%; Score 68; DB 4; Length 503;

Best Local Similarity 28.48; Pred. No. 1;
Matches 21; Conservative 18; Mismatches 29; Indels 6; Gaps 2

```

Oy 27 VMEETIHNTAMMRKSLSEASIVREKVALIKCELOSTINGSGGEAMVLYEWRSLSAEG 86
      | : : | | : | : : | : : | : : : | : : : |
Db 159 VLNQTLNKKQOMLENSL-STKKLERQIMMQSRDLQRLGRN---RALETRRLQALEAQH 212

```

```
QY      87 QAHVLALRDVEEQI 100  
         || : : | : : ||  
Db     213 QAQLNSLQEKREQL 226
```

RESULT 2
US-08-665-926-8

```

; GENERAL INFORMATION:
APPLICANT: Valenzuela et al.

```

```
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
ADDRESS: BOSTON
```

CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
COMPUTED: IBM PC

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: IIS/09/6665 036

```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Robert T. Robert

TELEPHONE: (014) 345-7400

```

; INFORMATION FOR SEQ ID NO:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 509 amino acids

```

```

;      TOPOLOGY:  linear
;      MOLECULE TYPE:  protein
US-08-665-926-8

```

Query Match	Score	DB	Length
Best local similarity	13.8%	68	509
	28.4%	Prod No. 1	

0y 27 VMEETIHYNTAMRASLEASTVREVALKICELQJMGSGEAMAVLVYENRSLAEG 86

QY 87 QAHVLRDVEOI 100
 || . . . ||
 || . . . ||

RESULT 3

```

; Sequence 10, Application US/08740223A
; Patent No. 6265564
;
; GENERAL INFORMATION:

```

TITLE OF INVENTION: Expressed Ligand - Vascular

```

; TITLE OF INVENTION: Intercellular Signalling Molecules
;
; NUMBER OF SEQUENCES: 28
;

```

ADDRESSEE: Regeneron Pharmaceuticals, Inc
STREET: 777 Old Saw Mill Road
CITY: Tarrytown

; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE

```

;
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.0
;

```

; FILING DATE: 25-OCT-1996
 ; CLASSIFICATION: 536

;; FILING DATE: 02-AUG-1996
;; ATTORNEY/AGENT INFORMATION:

REFERENCE/DOCKET NUMBER: REG 3333
TELECOMMUNICATION INFORMATION:

```

; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 nt

```

```
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```

; FEATURE: TIE ligand-3
; NAME/KEY: TIE ligand-3
; LOCATION: 1 E06

```

US-08-740-223A-10

Best Local Similarity 28.4%; Pred. NO. 1;
Matches 21; Conservative 18; Mismatches 29; Indels 6; Gaps 2;

```
Db      165 VLNQTLHMKTOMLENSL--STNKLEROMIQSRELOQRN---RALETRLQLALEAQH 218
```

```

      219 QAQLNSLOEKREQL 232
Db

```

RESULT 4
US-08-472-028A-4
SECURITY 4 ANALYST: JMS/08472028A

```

; GENERAL INFORMATION:
; APPLICANT: ward,
; ADDRESSEE: vol 1

```

1 TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
2
3 NUMBER OF SEQUENCES: 12
4
5 CORRESPONDENCE ADDRESS:

STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: NY

```
; ZIP: 10532
```



```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/012,705
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/013,612
:

```

; FILING DATE: 21-JUN-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Meigs, J. Timothy
 ; REGISTRATION NUMBER: 38,241
 ; REFERENCE/DOCKET NUMBER: CGC 1846

```
TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 508 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-808-323-4

Query Match      13.7%  Score 67.5; DB 3; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFSKDEILRAVM-----EETIHNTAMMRAS-----LEASTVREVLALIRCELOSIM 65
Db 366 HGFTLGLTFSSMPEPDRSPDVHLTYFTFGSRNOELAKAST--DELKQVTSDLQRL 423
QY 66 GSGSEMAVLVEWR 80
Db 424 GVEGEPVSVNHYYWR 438

RESULT 7
US-09-050-603A-4
: Sequence 4, Application US/09050603A
: Patent No. 6023012
: GENERAL INFORMATION:
: APPLICANT: Voltrath, Sandra
: APPLICANT: Johnson, Marie
: APPLICANT: Potter, Sharon
: APPLICANT: Ward, Eric
: APPLICANT: Helfetz, Peter
: TITLE OF INVENTION: DNA Molecules Encoding Plant
: TITLE OF INVENTION: Prototophyllinogen Oxidase
: NUMBER OF SEQUENCES: 37
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6023012artls Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/050,603A
: FILING DATE: 30-MAR-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/808,931
: FILING DATE: 28-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/012,705
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/013,612
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/020,003
: FILING DATE: 21-JUN-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Meigs, J. Timothy
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: CGC 1847
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (919) 541-8587
: TELEFAX: (919) 541-8689
```

```
INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 508 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-050-603A-4

Query Match      13.7%  Score 67.5; DB 3; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHFSKDEILRAVM-----EETIHNTAMMRAS-----LEASTVREVLALIRCELOSIM 65
Db 366 HGFTLGLTFSSMPEPDRSPDVHLTYFTFGSRNOELAKAST--DELKQVTSDLQRL 423
QY 66 GSGSEMAVLVEWR 80
Db 424 GVEGEPVSVNHYYWR 438

RESULT 8
US-09-102-420B-4
: Sequence 4, Application US/09102420B
: Patent No. 6084155
: GENERAL INFORMATION:
: APPLICANT: Voltrath, Sandra
: APPLICANT: Johnson, Marie
: APPLICANT: Ward, Eric
: APPLICANT: Helfetz, Peter
: TITLE OF INVENTION: HERBICIDE-TOLERANT PROTOPORPHYRINOGEN
: TITLE OF INVENTION: OXIDASE ("PROTOX")
: NUMBER OF SEQUENCES: 43
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 6084155artls Corporation
: STREET: 3054 Cornwallis Road
: CITY: Research Triangle Park
: STATE: NC
: COUNTRY: USA
: ZIP: 27709
: COMPUTER READABLE FORM:
: MEDIUM TYPE: floppy disk
: OPERATING SYSTEM: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/102,420B
: FILING DATE: 22-JUN-1998
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/059,164
: FILING DATE: 13-APR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 09/050,603
: FILING DATE: 30-MAR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/126,430
: FILING DATE: 11-MAR-1998
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/808,931
: FILING DATE: 28-FEB-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/012,705
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/013,612
: FILING DATE: 28-FEB-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/020,003
: FILING DATE: 21-JUN-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/472,028
```

US-09-071-296-4

```
QY      66  GSGSEAMAVLYEWR  80
          |  |  :  |  |  |
Db      424  GVEGEPVSVNHYYWR  438
```

US-09-071-296-4

```
RESULT 11
US-09-015-683-4
; Sequence 4, Application US/09015683
; Patent No. 6288306
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/015,683
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-015-683-4

Query Match      13.7%; Score 67.5; DB 4; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHKSKDEILRAVW----ETIHYNTPAMRAS---LEEASTVREVLALIRCELOSIM 65
| | | : : : | | : : : | | | : : : | | : : :
Db 366 HGKRTLTGTFSSMWFPPRSPSDVLYTTFTIGSGRNDELAKAST--DELKQVYTSDLQRL 423
| | | : : : | | : : : | | | : : : | | : : :
QY 66 GGSGEAMAVLYEWR 80
| | | : : : | | : : : | | | : : : | | : : :
Db 424 GVEGEPSVNVHYWR 438

RESULT 12
US-09-191-998-4
; Sequence 4, Application US/09191998
; Patent No. 6307129
; GENERAL INFORMATION:
; APPLICANT: Ward, Eric R
; APPLICANT: Volrath, Sandra
; TITLE OF INVENTION: Manipulation of Protoporphyrinogen
; TITLE OF INVENTION: Oxidase Enzyme Activity in Eukaryotic Organisms
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 7 Skyline Drive
; CITY: Hawthorne
; STATE: NY
; COUNTRY: USA
; ZIP: 10532
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,998
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,198
; FILING DATE: 16-JUN-94
; ATTORNEY/AGENT INFORMATION:
; NAME: Elmer, James Scott
; REGISTRATION NUMBER: 36,129
; REFERENCE/DOCKET NUMBER: CGC 1748/CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-541-8614
; TELEFAX: 919-541-8689
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 508 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-191-998-4

Query Match      13.7%; Score 67.5; DB 4; Length 508;
Best Local Similarity 26.7%; Pred. No. 1.2;
Matches 20; Conservative 16; Mismatches 28; Indels 11; Gaps 3;

QY 15 HHKSKDEILRAVW----ETIHYNTPAMRAS---LEEASTVREVLALIRCELOSIM 65
| | | : : : | | : : : | | | : : : | | : : :
Db 366 HGKRTLTGTFSSMWFPPRSPSDVLYTTFTIGSGRNDELAKAST--DELKQVYTSDLQRL 423
| | | : : : | | : : : | | | : : : | | : : :
QY 66 GGSGEAMAVLYEWR 80
| | | : : : | | : : : | | | : : : | | : : :
Db 424 GVEGEPSVNVHYWR 438

RESULT 13
US-08-449-933-2
; Sequence 2, Application US/08449933
; Patent No. 5859195
; GENERAL INFORMATION:
; APPLICANT: Collins, Francis S.
; APPLICANT: Wallace, Margaret R.
; APPLICANT: Marchuk, Douglas A.
; APPLICANT: Anderson, Lone B.
; APPLICANT: Gultman, David H.
; TITLE OF INVENTION: Neurofibromatosis Gene
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morrison & Foerster
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,933
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Konski, Antoinette F.
; REGISTRATION NUMBER: 34,202
; REFERENCE/DOCKET NUMBER: 20344-20553.10
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (415) 813-5600
 TELEFAX: (415) 494-0792
 TELEEX: 706141 MRSNEFERS SFO
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2818 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 HYPOTHEICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT: 17q11.2
 FEATURE:
 NAME/KEY: Cleavage-site
 LOCATION: group(583..586, 815..818, 2573..2576, 2810..2813)
 OTHER INFORMATION: /note= "Potential CAMP-dependent
 OTHER INFORMATION: protein kinase recognition sites"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 2549..2556
 OTHER INFORMATION: /note= "Potential tyrosine
 OTHER INFORMATION: phosphorylation site"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: group(1264, 1276, 1358, 1377, 1389, 1390, 1391,
 LOCATION: 1395, 1396, 1400, 1423, 1426, 1429, 1430)
 OTHER INFORMATION: /note= "Invariant residues within
 OTHER INFORMATION: most statistically significant regions of similarity among th
 OTHER INFORMATION: GAP family of proteins"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: group(1264..1290, 1345..1407, 1415..1430)
 OTHER INFORMATION: /note= "Most statistically
 OTHER INFORMATION: significant regions of similarity among the GAP family of
 OTHER INFORMATION: proteins"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 496
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence which shows an ATG methionine codon rather
 OTHER INFORMATION: than an ATA isoleucine codon"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 1183
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence. Shows an CTG leucine codon rather than th
 OTHER INFORMATION: previously published CTC"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 1555
 OTHER INFORMATION: /note= "At variance with previously
 OTHER INFORMATION: published sequence. Lacks an extra CAT histidine condon after
 OTHER INFORMATION: this residue"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: (2771-2772)
 OTHER INFORMATION: /note= "Position of an 18 amino
 OTHER INFORMATION: acid insertion(SEQ ID NO:10) representing an alternatively
 OTHER INFORMATION: spliced product"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: (1370-1371)
 OTHER INFORMATION: /note= "Position of a 21 amino acid
 OTHER INFORMATION: insertion representing an alternatively spliced product"
 FEATURE:
 NAME/KEY: Domain
 LOCATION: 1125..1537
 OTHER INFORMATION: /note= "NFI catalytic domain"
 FEATURE:

NAME/KEY: Modified-site
 LOCATION: 2746..2818
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the PstI-HindIII fragment designated pMAL.B3A"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 65..371
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the HpaI-PstI fragment designated pMAL.HF3A.P"
 FEATURE:
 NAME/KEY: Modified-site
 LOCATION: 65..1240
 OTHER INFORMATION: /note= "Corresponding amino acids
 OTHER INFORMATION: for the HpaI-XhoI fragment designated pMAL.HF3A.X"
 PUBLICATION INFORMATION:
 AUTHORS: Wallace, M.R. et al.
 TITLE: Type I Neurofibromatosis Gene: Correction
 JOURNAL: Science
 VOLUME: 250
 ISSUE: 12/21/90
 PAGES: 1749-
 DATE: 12/21-1990
 RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818
 PUBLICATION INFORMATION:
 AUTHORS: Wallace, M.R. et al.
 TITLE: Type I Neurofibromatosis Gene: Identification
 TITLE: of a Large Transcript in Three NFI Patients
 JOURNAL: Science
 VOLUME: 249
 ISSUE: 07/13/90
 PAGES: 181-186
 DATE: 07/13-1990
 RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818
 US-08-449-933-2

Query Match 13.6%; Score 67; DB 2; Length 2818;
 Best Local Similarity 26.5%; Pred. No. 17;
 Matches 26; Conservative 18; Mismatches 32; Indels 22; Gaps 4;
 QY 12 STFHFKSKDELRAVMEITH-----YNTAMNRASLEE-----ASTVERVLALIRCL 61
 DB 708 SCFRIHCEADIRCAVEDEVSHLLPNYTFMEFASVSNMSTGRALOKRYMALLR-RI 766
 QY 62 OSIMGSGEAMAVLYEY-----RSLASGQA 88
 DB 767 EHPYAGNTEAMEDTHAKMEQATKLLINPKAKMEDQA 804

RESULT 14
 US-07-966-049A-2
 Sequence 2, Application US/07966049A
 Patent No. 6238861
 GENERAL INFORMATION:
 APPLICANT: Collins, Francis S.
 APPLICANT: Wallace, Margaret R.
 APPLICANT: Marchuk, Douglas A.
 APPLICANT: Anderson, Lone B.
 APPLICANT: Gultman, David H.
 TITLE OF INVENTION: Neurofibromatosis Gene
 NUMBER OF SEQUENCES: 14
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Morrison & Foerster
 STREET: 755 Page Mill Road
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304-1018
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/966,049A
FILING DATE: 02-AUG-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Komski, Antoinette F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-20553.20
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792
TELEX: 706141 MRSNFOERS SFO
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2818 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
POSITION IN GENOME:
CHROMOSOME/SEGMENT: 17q11.2
FEATURE:
NAME/KEY: Cleavage-site
LOCATION: group(1264..586, 815..818, 2573..2576, 2810..2813)
OTHER INFORMATION: /note="Potential CAMP-dependent"
OTHER INFORMATION: protein kinase recognition sites"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 2549..2556
OTHER INFORMATION: /note="Potential tyrosine phosphorylation site"
FEATURE:
NAME/KEY: Modified-site
LOCATION: group(1264, 1276, 1358, 1377, 1389, 1390, 1391,
LOCATION: 1395, 1396, 1400, 1423, 1426, 1429, 1430)
OTHER INFORMATION: /note="Invariant residues within"
OTHER INFORMATION: most statistically significant regions of similarity among th
OTHER INFORMATION: GAP family of proteins"
FEATURE:
NAME/KEY: Modified-site
LOCATION: group(1264..1290, 1345..1407, 1415..1430)
OTHER INFORMATION: /note="Most statistically
OTHER INFORMATION: significant regions of similarity among the GAP family of
OTHER INFORMATION: proteins"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 466
OTHER INFORMATION: /note="At variance with previously
OTHER INFORMATION: published sequence which shows an ATG methionine codon rather
OTHER INFORMATION: than an ATA isoleucine codon"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1183
OTHER INFORMATION: /note="At variance with previously
OTHER INFORMATION: published sequence. Shows an C TG leucine codon rather than th
OTHER INFORMATION: previously published CTC"
FEATURE:
NAME/KEY: Modified-site
LOCATION: 1555
OTHER INFORMATION: /note="At variance with previously
OTHER INFORMATION: published sequence. Lacks an extra CAT histidine condon after
OTHER INFORMATION: this residue"
FEATURE:
NAME/KEY: Modified-site
LOCATION: (2771~2772)
OTHER INFORMATION: /note="Position of an 18 amino
OTHER INFORMATION: acid insertion(SEQ ID NO:10) representing an alternatively
OTHER INFORMATION: spliced product"
FEATURE:

```

```

? NAME/KEY: Modified-site
? LOCATION: (1370-1371)
? OTHER INFORMATION: /note= "Position of a 21 amino acid
? OTHER INFORMATION: Insertion representing an alternatively spliced product"
? FEATURE:
? NAME/KEY: Domain
? LOCATION: 1125..1537
? OTHER INFORMATION: for the PstI-HindIII fragment designated pMAL.B3A"
? FEATURE:
? NAME/KEY: Modified-site
? LOCATION: 65..371
? OTHER INFORMATION: /note= "Corresponding amino acids
? OTHER INFORMATION: for the HpaI-PstI fragment designated pMAL.HF3A.P"
? FEATURE:
? NAME/KEY: Modified-site
? LOCATION: 65..1240
? OTHER INFORMATION: /note= "Corresponding amino acids
? OTHER INFORMATION: for the HpaI-XhoI fragment designated pMAL.HF3A.X"
? PUBLICATION INFORMATION:
? AUTHORS: Wallace, M.R. et al.
? TITLE: Type I Neurofibromatosis Gene: Correction
? JOURNAL: Science
? VOLUME: 250
? ISSUE: 12/21/90
? PAGES: 1749-
? DATE: 12/21-1990
? RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818
? PUBLICATION INFORMATION:
? AUTHORS: Wallace, M.R. et al.
? TITLE: Type I Neurofibromatosis Gene: Identification
? TITLE: of a Large Transcript In Three NFI Patients
? JOURNAL: Science
? VOLUME: 249
? ISSUE: 07/13/90
? PAGES: 181-186
? DATE: 07/13-1990
? RELEVANT RESIDUES IN SEQ ID NO: 2: FROM 1 TO 2818
US-07-966-049A-2

Query Match          13.6%; Score 67.; DB 4; Length 2818;
Best Local Similarity 26.5%; Pred. No. 17;
Matches 26; Conservative 18; Mismatches 32; Indels 22; Gaps 4;

QY 12 SIFHHKSKDILRLAVYMETH-----YNPMMASLSE-----ASYRRRLALIRCEL 61
   |||||.....| | | | | | | | | | | | | | | | | | | | | | | | | |
DB 708 SCFRLLCEADIKCAVDVSVHNLPLNPTFMERASVSNNMSTGRALLQRKVALLR-RI 766
               : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 62 QSIMGSGEAMAVLYEW-----RSLAEGQA 88
   : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 767 EHPAGNTAEAWEDTHAKWEQATKLILNVPRKKMEDGA 804
               : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 15
US-08-882-704A-2
Sequence 2, Application US/08882704A
Patent No. 5879906
GENERAL INFORMATION:
APPLICANT: Jefferson, Richard A.
APPLICANT: Wilson, Katherine J.
TITLE OF INVENTION: GLUCURONIDE REPRESSORS AND USES THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
```

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/882,704A
FILING DATE: 25-JUN-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: No. 5879906tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 190106.404
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 195 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: Linear
US-08-882-704A-2

Query Match 13.3%; Score 65.5; DB 2; Length 195;
Best Local Similarity 26.4%; Pred. No. 0.55;
Matches 23; Conservative 17; Mismatches 34; Indels 13; Gaps 3;

QY 2 LASAVGIQSGSIFHHKSKDELIRAVMEETIHNTAMRASLEASTVRERYLALIRCEL 61
DB 36 ICKSCAISPGLTHTHHSKALIQATILLODERALAREPRELEGIFVDYMY----- 87
QY 62 OSIMGSGCEAMA--VLVYEMRSLSAAG 86
DB 88 ESIVSLTHEAFGQALVYE---IMAE 111

Search completed: March 4, 2002, 20:24:35
Job time: 83 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:23:32 ; Search time 19 Seconds
(without alignments)
400.919 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIQGSIFHFKSK.....SLSAEGAHVLAIRDYEQI 100

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	429	87.0	198	2	C83286
2	110.5	22.4	190	2	S38906
3	110.5	22.4	197	2	F75281
4	94.5	19.2	200	2	C70604
5	92.5	18.8	194	2	F69985
6	92.5	18.8	215	2	F84019
7	89	18.1	192	2	C70487
8	86	17.4	195	2	F84037
9	85	17.2	197	2	G82973
10	84.5	17.1	215	2	T36800
11	84	17.0	206	2	T35415
12	82.5	16.7	223	2	T35424
13	81	16.4	190	2	D70027
14	80	16.2	186	2	B83491
15	80	16.2	212	2	H83274
16	79	16.0	209	2	T36702
17	77.5	15.7	187	2	T36803
18	77.5	15.7	194	2	A83371
19	77	15.6	198	2	F70328
20	77	15.6	216	2	A81156
21	77	15.6	251	2	T36792
22	76.5	15.5	216	2	B81950
23	76.5	15.5	217	2	T35787
24	76	15.4	195	2	T36378
25	76	15.4	213	2	T36803
26	75.5	15.3	195	2	C85524
27	75.5	15.3	195	2	S10899
28	75.5	15.3	209	2	T34970
29	75	15.2	193	2	D83417

30	74	15.0	188	2	B70736	hypothetical prote
31	74	15.0	225	2	C70500	probable transcrip
32	73.5	14.9	172	2	G84416	hypothetical prote
33	73.5	14.9	210	2	E83471	probable transcrip
34	73.5	14.9	210	2	F83393	probable transcrip
35	73	14.8	176	2	C55208	socA3 protein - My
36	72.5	14.7	186	2	C83042	probable transcrip
37	72.5	14.7	213	2	F70946	probable regulato
38	72	14.6	197	2	A83694	hypothetical prote
39	72	14.6	204	2	F69105	conserved hypotet
40	72	14.6	213	2	F82357	transcription regu
41	71.5	14.5	199	2	B85772	hypothetical prote
42	71.5	14.5	201	2	G70849	probable transcrip
43	71.5	14.5	232	2	A57507	virginiae butanol
44	71	14.4	185	2	D83703	transcription regu
45	71	14.4	200	2	E82203	transcription regu

ALIGNMENTS

RESULT 1
C83286
probable transcription regulator PA2885 [Imported] - Pseudomonas aeruginosa (strain P
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83286
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Latig, K.; L
.; Lofy, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: C83286
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <STO>
A:Cross-references: GB:AE004714; GB:AE004091; NID:g9948965; PIDN:AA06273.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2885

Query Match 87.0%; Score 429; DB 2; Length 198;
Best Local Similarity 84.0%; Pred. No. 1,8e-38;
Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;
QY 1 DLASAVGIQGSIFHFKSKDEILRAVMEETIHYNTAMKRSLEASTYRERYALRICE 60
|||||
Db 40 DLASAVGIQGSIFHFKSKDEILRAVMEETIHYNTAMKRSLEASTYRERYALRICE 99
|||||
QY 61 LOSIMGSGEAMAVLYEWRSLSAEGAHVLAIRDYEQI 100
|||||
Db 100 LOSIMGSGEAMAVLYEWRSLSAEGAHVLAIRDYEQI 139
|||||
RESULT 2
S38906
hypothetical protein 4 - Clostridium pasteurianum
C:Species: Clostridium pasteurianum
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 08-Oct-1999
R:Meyer, J.
submitted to the EMBL Data Library, November 1993
A:Description: Sequence of 6764 bp EcoRI-Sau3A fragment of Clostridium pasteurianum g
ry M62754).
A:Reference number: S38903
A:Accession: S38906
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-190 <MEY>
A:Cross-references: EMBL:Z28353; NID:g431946; PIDN:CAA82211.1; PID:g431950
C:Superfamily: Bacillus subtilis probable transcription regulator yrhI

Query Match	22.4%	Score 110.5;	DB 2;	Length 197;
Best Local Similarity	27.5%	Pred. No. 0.00017;		
Matches	28;	Conservative	23;	Mismatches 42;
			Indels	9;
			Gaps	2;

QY	1	DLASVAGTSSIFPHKSKDEL	----	RAVMEETHYNTAMRASLEPASTVREYVAL	56
		: :	: :	: :	
DB	32	DLAQGLQGGSVLVNIHSGKEELLEIVTIRGASQCFDEALFSLRDVNLPADEKLREAMPFRH			91
		: :	: :	: :	
QY	57	IRCELOSTIGGSGEAMAVLVYEWMSLRSLAGCAHYVALDYVE			98
		: : :	: :	: :	
DB	92	IQ-----VVADNMDSATVFFHEMKHLSLSEPYAAQVVAMPDDTTD			128

RESULT 5
F69985
transcription regulator TetR/AcrR family homolog ysaA - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: F69985
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bern
C.: Ehrlich, S.D.; Emmerson, P.T.; Entlian, K.D.; Errington, J.; Fabret, C.; Ferrari,
A.: Ehrlich, S.D.; Emmerson, P.T.; Entlian, K.D.; Errington, J.; Fabret, C.; Ferrari,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
lech, Y.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holseptel, S.; Hosono, S.; Hullo,
Koester, P.; Koningsstein, G.; Krogh, A.; Kunano, M.; Kurita, K.; Lapidus, A.; Lardine
A.:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y.:M.; Ogawa, K.; Oglwata, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl
A:Authors: Schleich, S.; Schroeder, R.; Scollone, F.; Sekiguchi, J.; Sekowska, A.; S
akench, M.; Tamakoshi, A.; Tanaka, T.; Terpestä, P.; Toignon, A.; Tosaio, V.; Uchiy
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, Y.; Dancin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis
#:Reference number: A69580; MUID:98044033

RESULT 6
F84019
Transcription regulator (TelR/AcrR family) BH2958 [imported] - *Bacillus halodurans* (S)
C:Species: *Bacillus halodurans*
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: F84019
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* s
A:Reference number: A83650; MUID:20263314
A:Accession: F84019
A:Status: preliminary
A:Molecule type: DNA

46 DIAGRAGMSPALYIHYKTKELLHRISRICHTRAVALLRSAQGECSAERLADAVSSF 105

Tue Mar 5 07:57:18 2002

Db 91 AARLFQDDSIGCLMGVAVD---ASYGRSEIMA 120

RESULT 15

H83274

probable transcription regulator PA2957 [imported] - Pseudomonas aeruginosa (strain PA01
C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: H83274

R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; Lim,

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337

A:Accession: H83274

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-212 <STO>

A:Cross-references: GB:AE004722; GB:AE004091; NID:g9949054; PIDN:AG06345.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA2957

Query Match 16.2%; Score 80; DB 2; Length 212;

Best Local Similarity 27.4%; Pred. NO. 0.32;

Matches 29; Conservative 12; Mismatches 35; Indels 30; Gaps 3;

QY 2 LASAVGIQSGSIFHHFKSKDELIRAVM-----EETIHYNTAMMRA----- 41

Db 39 IADAVGIGKGTIKYKFKAEIYLRLMDYERDLALFHSEDEVARDKEALSRAVFEFRMR 98

QY 42 -----SLEASTVRERYLALIRCELOSIMGSGEAMAVLYE 78

Db 99 DPORYRLFDRLERKVVTSQVPEWE-ELHKIRASNFERTOLIKE 143

Search completed: March 4, 2002, 20:25:06
Job time: 94 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:27:17 ; Search time 10.34 Seconds

(without alignments)
354.592 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIGSGSIFHFHRSK.....SLSAEQAVLALRDVYEQI 100

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.5	16.5	203	1	BETI_RHIME
2	75.5	15.3	195	1	BETI_ECOLI
3	74	15.0	188	1	XY05_MYCTU
4	72	14.6	2090	1	HFC1_MESAU
5	70.5	14.3	215	1	YIIC_ECOLI
6	70	14.2	187	1	Y893_ECOLI
7	70	14.2	676	1	HS7C_TRYB
8	69.5	14.1	452	1	NH34_CAEBL
9	68.5	13.9	267	1	Y617_SYNY3
10	68.5	13.9	289	1	YERO_BACSU
11	68	13.8	391	1	IF43_NICPL
12	68	13.8	4568	1	DYHC_CAEBL
13	67.5	13.7	226	1	TCMR_STRGA
14	67.5	13.7	1814	1	TSC2_MOUSE
15	67	13.6	210	1	MTRR_NEIGO
16	67	13.6	215	1	ACRR_ECOLI
17	66.5	13.5	199	1	YOHM_ECOLI
18	66.5	13.5	917	1	Y6J3_YEAST
19	66	13.4	218	1	TKR_HAELN
20	66	13.4	223	1	YWCQ_BACSU
21	65.5	13.3	196	1	UIDR_ECOLI
22	65.5	13.3	949	1	TSC2_RHIME
23	65.5	13.3	1809	1	TSC2_RAT
24	65	13.2	198	1	TKR_ECOLI
25	65	13.2	1529	1	PDRF_YEAST
26	64.5	13.1	471	1	VG20_BPP22
27	64	13.0	275	1	PORI_PEA
28	64	13.0	3746	1	ACVS_PENCH
29	64	13.0	3791	1	ACVT_PENCH
30	63.5	12.9	1807	1	TSC2_HUMAN
31	63	12.8	338	1	LAMP_CHICK
32	63	12.8	885	1	MVP_RAT
33	63	12.8	2820	1	NF1_RAT

34	63	12.8	2839	1	NF1_HUMAN	P21359 homo sapien
35	62.5	12.7	455	1	VPN_HAELN	P71389 haemophilus
36	62	12.6	174	1	NADM_ARCFU	O27968 archaeoglob
37	62	12.6	219	1	TER3_ECOLI	P03093 escherichia
38	62	12.6	358	1	HRCA_CAUCR	P54305 caulobacter
39	62	12.6	428	1	GFAP_BOVIN	Q28115 bos taurus
40	62	12.6	437	1	MURD_HAELN	P45063 haemophilus
41	62	12.6	461	1	VIME_ONCMY	P48674 oncorhynchus
42	62	12.6	558	1	ALKJ_PSEOL	O00593 pseudomonas
43	61.5	12.5	262	1	OCCP_AGRU	P51117 agrobacteri
44	61.5	12.5	3255	1	POLG_LMWE	P89876 1 genome po
45	61	12.4	210	1	YCFQ_ECOLI	P75952 escherichia

ALIGNMENTS

RESULT 1	ID	BETI_RHIME	STANDARD:	PRT:	203 AA.
AC	O69786:	BETI_RHIME			
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	REGULATORY PROTEIN BETI.				
CN	BETI.				
OS	Rhizobium meliloti (Sinorhizobium meliloti).				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.				
OX	Rhizobiaceae; Sinorhizobium.				
OX	NCBI_TaxID=382;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-102F34;				
RX	MEDLINE-96409668; PubMed-9736747;				
RA	Oesteras M., Boncompagni E., Vincent N., Poggi M.-C., Le Rudulier D.;				
RT	"Presence of a gene encoding choline sulfatase in Sinorhizobium				
RT	meliloti bet operon: choline-O-sulfate is metabolized into glycine				
RT	betaine."				
RL	Proc. Natl. Acad. Sci. U.S.A. 95:11394-11399(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-1021;				
RA	Gallbert F., Capela D., Hubler-Barloy F., Gattus M., Batut J.,				
RA	Boisard P., Gouzy J., Kahn D., Thebault P., Goffeau A.,				
RA	Pumelle B., Pohl T., Bothe G., Schneider S., Portetelle D.,				
RL	Vandenbol M., Puehler A., Becker A., Weidner S.;				
CC	Submitted (MAR-2000) to the SWISS-PROT data bank.				
CC	- FUNCTION: REPRESSOR INVOLVED IN CHOLINE REGULATION OF THE BET				
CC	GENES (BY SIMILARITY).				
CC	- PATHWAY: OSMOREGULATORY CHOLINE-GLYCINE BETAINES PATHWAY.				
CC	- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL				
CC	REGULATORS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U39940; AAC13370.1; -				
DR	InterPro: IPR001647; HTH_Tetr.				
DR	Plan: PF00440; tetr. 1.				
DR	PROSITE: PS01081; HTH_TETR_FAMILY; FALSE_NEG.				
KW	Transcription regulation; DNA-binding; Repressor; Trans-acting factor.				
FT	DNA_BIND 31 50 H-T-H MOTIF (BY SIMILARITY).				
SO	SEQUENCE 203 AA; 22039 MW; 749F140947339A13 CRC64;				

Query Match 16.5%; Score 81.5; DB 1; Length 203;
Best Local Similarity 26.4%; Pred. No. 0.11;
Matches 23; Conservative 18; Mismatches 41; Indels 5; Gaps 1;

```

OY      1  DLASVAGSIOSSIFHHFHSKDIPLAAWMEETHYNTAMTASLEASTYREVALLICE 60
Db      34  ELATRACGSPALAHNYHFFOSKEOLLATATIRSLGLRDAVAAKAAATIPREXVALIKRS 93
OY      61  LOSINGSGSEAMAVLVEMRSLASGQ 87
Db      94  FR----ADQFAPETVAAWLAIFYSEAQ 115

RESULT      2
BETI_ECOLI STANDARD; PRT; 195 AA.
AC      P17446;
DT      01-AUG-1990 (Rel. 15, Created)
DT      01-AUG-1990 (Rel. 15, Last sequence update)
DT      20-AUG-2001 (Rel. 40, Last annotation update)
DE      REGULATORY PROTEIN BETI.
GN      BETI OR B0313.
OS      Escherichia coli.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia.
OX      NCBI_TaxID=562;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12;
RX      MEDLINE=92065800; PubMed=1956285;
RA      Lemark T., Kaasen E., Eshoo M.W., Falkenberg P., McDougall J.,
RA      Strom A.R.;
RT      "DNA sequence and analysis of the bet genes encoding the
RT      osmoregulatory choline-glycine betaine pathway of Escherichia coli.";
RL      Mol. Microbiol. 5:1049-1064(1991).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / MC1655;
RX      MEDLINE=97426617; PubMed=9278603;
RA      Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA      Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA      Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA      Mau B., Shao Y.;
RT      "The complete genome sequence of Escherichia coli K-12.";
RL      Science 277:1453-1474(1997).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      Duncanson M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA      Federapfel M., Hyman R., Kallman S., Komp C., Kirdi O., Lew H.,
RA      Lin D., Nemeth A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL      Submitted (OCT-1996) to the EMBL/Genbank/DBJ databases.
CC      -1- FUNCTION: REPRESSOR INVOLVED IN CHOLINE REGULATION OF THE BET
CC      GENES.
CC      -1- PATHWAY: OSMOREGULATORY CHOLINE-GLYCINE BETAINE PATHWAY.
CC      -1- SIMILARITY: BELONGS TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL
CC      REGULATORS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on ways
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X52905; CAA37091.1; -
DR      EMBL; AE000138; AAC73416.1; -
DR      EMBL; U73857; AAB18039.1; ALT_INIT.
DR      PIR; S10899; S10899.
DR      PIR; S15180; S15180.
DR      EcoGene; EG1011; BetI.
DR      InterPro; IPR001647; HTH_Tetr.
DR      Pfam; PF00440; tetr. 1.
DR      PROSITE; PS01081; HTH_TETR_FAMILY; 1.
KW      Transcription regulation; DNA-binding; Repressor; Trans-acting factor;
KW      Complete proteome.

```

```

FT   DNA_BIND      31          50          H-T-H MOTIF (BY SIMILARITY).
SQ   SEQUENCE    195 AA;  21815 MW;  B52AV77B3A605E354 CRC64;
Query Match      15.3%; Score 75.5; DB 1; Length 195;
Best Local Similarity 28.6%; Pred. No. 0.43;
Matches 26; Conservative 14; Mismatches 30; Indels 21; Gaps 4;

OY  2 LASAAGIOSGSPHFHKSKDELRAVMEETIHYNNAMRASLEASATREHYAL- - - I 57
DB  35 IARRAGVSTGLTISHYFRKRNGLILEATMDT- - - TSOLR- - - DAVLNRLHALPQSA 84
OY  58 RCELOISMG- - - SGEMAAVLVEWRS 81
DB  85 EORLQAIVGGNFDETOVSAAAKMALAEPMAS 115

RESULT      3
YY05_MYCTU STANDARD; PRT; 188 AA.
AC O50720;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHEICAL 20.6 KDA PROTEIN RV3405C.
GN RV3405C OR MT3513 OR MTCY78.23.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Acidimycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felkell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsbly T., Jagielski K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the genome of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Allard D., Eisen J.A., Carpenter C., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayan L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Knout H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
[3]
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; Z77165; CAB01018.1; -.
DR EMBL; AE007157; AAK47851.1; -.
DR TIGR; MT3513; -.
DR Tuberculist; RV3405C; -.
DR InterPro; IPRO01647; HTH_Tetr.
DR Pfam; PF00440; tetr_1
DR PRINTS; PR00455; HTHHTTR.
KW Hypothetical protein; Complete proteome.

```


DT 30-MAY-2000 (Rel. 39, last annotation update)
DE NUCLEAR HORMONE RECEPTOR FAMILY MEMBER NHR-34
GN NHR-34 OR F58G6.5.
OS *Caenorhabditis elegans*.

OC Eukaryota: Metazoa: Nematoda; Chromodorea; Rhabditida; Rhabditoidea;
OC Rhabditidae: Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Lloyd C.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ORPHAN NUCLEAR RECEPTOR.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68217; CA92467.1; -
DR HSSP: P20393; IAGY.
DR Wormpep: F5866.5; CE03432.
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001628; zf-C4.
DR Pfam: PF00104; hormone_rec.1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00047; STROIDFINGER.
DR SMART: SM00430; HOL1; 1.
DR SMART: SM00399; znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor: Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
FT DNA_BIND 19 84 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 19 39 C4-TYPE.
FT ZN_FING 55 79 C4-TYPE.
SO SEQUENCE 452 AA; 51885 MW; 1A08677726C391CF CRC64;

Query Match 14.1%; Score 69.5; DB 1; Length 452;
Best Local Similarity 26.1%; Pred. No. 5;
Matches 24; Conservative 12; Mismatches 31; Indels 25; Gaps 3;

OY 6 VGIOGSIFFHFKSKDELIRAVMEETIHYNTAMRASLEASTYREYRLALIRCLQSIM 65
Db 301 VALNGRCNHFRIISEHL--MEDLV--IPREMDMEGEFYLKALILFRAH----- 349
OY 66 GSGEAMAVLYEWRSLSAEGCAHVLAIRDYV 97
Db 350 -----RRLSEGRRAHKIRVDKY 367

RESULT 9
Y617-SYNY3
ID Y617-SYNY3 STANDARD: PRT; 267 AA.
AC 055707;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 28.9 KDA PROTEIN SL0617.
GN SL0617.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria: Cyanobacteria: Chroococcales: Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE:96127529; PubMed:8590279;
RA Kaneo T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome."
RT DNA Res. 2:153-166(1995).

CC -!- SIMILARITY: BELONGS TO THE PSPA/IM30 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D64002; BAAL0345.1; -
DR Hypothetical protein; Coiled coil; Complete proteome.
KW DOMAIN 26 156 COILED COIL (POTENTIAL).
SO SEQUENCE 267 AA; 28905 MW; 5DDE309FFB0FF1A6 CRC64;

Query Match 13.9%; Score 68.5; DB 1; Length 267;
Best Local Similarity 30.4%; Pred. No. 3.5;
Matches 17; Conservative 9; Mismatches 25; Indels 5; Gaps 1;

OY 28 MEETIHYNTAMRASLEASTYREYRLALIR-----CELOSTMGSGEAMAVLYE 78
Db 114 MSENLRNMLALEAKISEAKTKKNMLQARAKANAELQDTLGLGTSATSATSE 169

RESULT 10
YERO_BACSU
ID YERO_BACSU STANDARD: PRT; 289 AA.
AC 031500;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL TRANSCRIPTIONAL REGULATOR YERO.
GN YERO.
OS Bacillus subtilis.
OC Bacteria: Firmicutes: Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z99107; CAB12490.1; -
DR Subtilist: BG12841; YERO.
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PRINTS: PR00455; HTHTEPR.
DR PROSITE: PS01081; HTH_TETR_FAMILY; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Complete proteome.
FT DNA_BIND 25 44 H-T-H MOTIF (POTENTIAL).
SO SEQUENCE 289 AA; 32846 MW; 540697223084DCBB CRC64;

Query Match 13.9%; Score 68.5; DB 1; Length 289;
Best Local Similarity 35.6%; Pred. No. 3.8;
Matches 16; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

OY 1 ELASAVGIQSGIFHFKSKDELIRAVMEETIHYNTAMRASLEAE 45
Db 28 ELASEGTSKGAFLYLFHFKSKALLLSACEYIYIGSMNKKK-NIEE 71

```

RESULT 11
ID IF43_NICPL STANDARD; PRT; 391 AA.
AC P41380;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EUKARYOTIC INITIATION FACTOR 4A-3 (EIF4A-3).
OS Nicotiana glauca (leadwort-leaved tobacco).
OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta:
OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots:
OC Asteridae: euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4092;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92051287; PubMed=1719476;
RA Owttrim G.W., Hofmann S., Kuhlmeier C.;
RT "Divergent genes for translation initiation factor eIF-4A are
RL Nucleic Acids Res. 19:5491-5496(1991).
CC -1- FUNCTION: EIF-4A IS BOTH A SUBUNIT OF A HIGH MOLECULAR WEIGHT
CC PROTEIN COMPLEX INVOLVED IN CAP RECOGNITION AND IS REQUIRED AS A
CC SINGLE POLYPEPTIDE CHAIN FOR MRNA BINDING TO RIBOSOME. IT IS AN
CC ATP-DEPENDENT UNWINDING ACTIVITY (HELICASE).
CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X61206; CAA43514.1; -.
CC InterPro: IPR001410; DEAD.
CC InterPro: IPR000629; DEAD_ATP_helcse.
CC InterPro: IPR001650; Helicase_C.
CC pfam: PF00271; helicase_C.1.
CC SMART: SM00487; DEXDC; 1.
CC SMART: SM00490; HELIC; 1.
CC PROSITE: PS00039; DEAD_ATP_HELICASE; FALSE_NEG.
CC Initiation factor: Protein biosynthesis: ATP-binding; RNA-binding;
CC DNA-binding; Helicase; Multigene family.
CC NP_BIND 62 69 ATP (BY SIMILARITY).
CC SITE 167 170 DEAD BOX.
CC SEQUENCE 391 AA; 44194 MW; 92CD2721916BCB1E CRC64;

Query Match 13.8%; Score 68; DB 1; Length 391;
Best Local Similarity 20.8%; Pred. No. 6.1;
Matches 21; Conservative 27; Mismatches 39; Indels 14; Gaps 2;

```

```

DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
OS Caenorhabditis elegans.
OC Eukaryota: Metazoa: Nematoda: Chromadorea: Rhabditida: Rhabditioidea;
OC Rhabditidae: Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2; PubMed=8674131;
RA Lye R.J., Wilson R.K., Waterston R.H.;
RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
RL nematode Caenorhabditis elegans."
RL Cell Motil. Cytoskeleton 32:26-36(1995).
CC -1- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -1- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: L33260; AAC37251.1; -.
CC InterPro: IPR003593; AAA.
CC SMART: SM00382; AAA; 1.
CC Motor protein: Microtubules; Dynein; ATP-binding; Coiled coil.
FT DOMAIN 587 652 COILED COIL (POTENTIAL).
FT DOMAIN 814 844 COILED COIL (POTENTIAL).
FT DOMAIN 1241 1274 COILED COIL (POTENTIAL).
FT DOMAIN 1324 1340 COILED COIL (POTENTIAL).
FT DOMAIN 1559 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1964 1992 COILED COIL (POTENTIAL).
FT DOMAIN 3132 3229 COILED COIL (POTENTIAL).
FT DOMAIN 3339 3432 COILED COIL (POTENTIAL).
FT DOMAIN 3707 3739 COILED COIL (POTENTIAL).
FT DOMAIN 4359 4386 COILED COIL (POTENTIAL).
FT NP_BIND 1865 1872 ATP (POTENTIAL).
FT NP_BIND 2163 2170 ATP (POTENTIAL).
FT NP_BIND 2537 2544 ATP (POTENTIAL).
FT NP_BIND 2880 2887 ATP (POTENTIAL).
CC SEQUENCE 4568 AA; 521568 MW; 028E52684F381676 CRC64;

Query Match 13.8%; Score 68; DB 1; Length 4568;
Best Local Similarity 27.7%; Pred. No. 1.1e+02;
Matches 26; Conservative 9; Mismatches 19; Indels 40; Gaps 4;

```

```

OS Streptomyces glaucescens.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomyces; Streptomyces; Streptomyces.
OX NCBI_TaxID=1907;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-GLA.0;
RX MEDLINE-92276347; PubMed=1592819;
RA Guilfoile P.G., Hutchinson C.R.;
RT "Sequence and transcriptional analysis of the Streptomyces
RT glaucescens tcmA tetracycline C resistance and repressor gene
RT loc1."
RL J. Bacteriol. 174:3651-3658(1992).
RN [2]
RP FUNCTION.
RX MEDLINE-92276348; PubMed=1592820;
RA Guilfoile P.G., Hutchinson C.R.;
RT "The Streptomyces glaucescens tcmA protein represses transcription of
RT the divergently oriented tcmA and tcmA genes by binding to an
RT intergenic operator region."
RL J. Bacteriol. 174:3659-3666(1992).
CC -1- FUNCTION: REPRESSES TRANSCRIPTION OF THE DIVERGENTLY ORIENTED TCMR
CC AND TCMR (TETRACYCLINE C RESISTANCE AND EXPORT) GENES BY BINDING
CC TO AN INTERGENIC OPERATOR REGION. THIS BINDING IS INHIBITED BY
CC TETRACYCLINE C.
CC -1- PATHWAY: BIOSYNTHESIS OF THE POLYKETIDE ANTIBIOTIC TETRACYCLINE
CC C.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M80674; AAA67508.1; ALT_SEQ.
DR PIR: B41901; B41901.
DR PIR: S27686; S27686.
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PRINTS: PR00455; HTH_Tetr.
DR PROSITE: PS01081; HTH_Tetr_FAMILY.1.
DR Antibiotic biosynthesis; Transcription regulation; DNA-binding;
DR Repressor.
DR DNA_BIND 49 68 H-T-H MOTIF (POTENTIAL).
DR SEQUENCE 226 AA; 25376 MW; 389C426A5054060C CRC64;

```

```

OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS A-G).
RC TISSUE-Heart;
RX MEDLINE-96258425; PubMed=8777431;
RA Kim K.R., Pajak L., Wang H., Field L.J.;
RT "Cloning, developmental expression, and evidence for alternative
RT splicing of the murine tuberous sclerosis (TSC2) gene product."
RL Cell. Mol. Biol. Res. 41:515-526(1995).
RN [2]
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE-96430093; PubMed=8833243;
RA Olsson P.G., Schofield J.N., Edwards Y.H., Frischauf A.M.;
RT "Expression and differential splicing of the mouse TSC2 homolog."
RL Mamm. Genome 7:212-215(1996).
RN [3]
RP SEQUENCE OF 1-199 FROM N.A.
RC STRAIN-BALB/C; TISSUE-Leukocyte;
RX MEDLINE-98417643; PubMed=9743625;
RA Sarker A.H., Ikeda S., Nakano H., Terato H., Ide H., Imai K.,
RA Akiyama K., Tsutsui K., Bo Z., Kubo K., Yamamoto K., Yasui A.,
RA Yoshida M.C., Seki S.;
RT "Cloning and characterization of a mouse homologue (mthl1) of
RT Escherichia coli endonuclease III."
RL J. Mol. Biol. 282:761-774(1998).
RN [4]
RP SEQUENCE OF 119-1805 FROM N.A.
RX MEDLINE-20051947; PubMed=10584558;
RA Kleymenova E.V., Declue J.E., Walker C.L.;
RT "Genetic variants of the tuberous sclerosis 2 tumour suppressor gene
RT in mouse haplotypes."
RL Genet. Res. 74:139-144(1999).
CC -1- FUNCTION: IMPLICATED AS A TUMOR SUPPRESSOR. MAY HAVE A FUNCTION IN
CC VESICULAR TRANSPORT, BUT MAY ALSO PLAY A ROLE IN THE REGULATION OF
CC CELL GROWTH ARREST AND IN THE REGULATION OF TRANSCRIPTION MEDIATED
CC BY STEROID RECEPTORS. INTERACTION BETWEEN HAMMARTIN AND TUBERIN MAY
CC FACILITATE VESICULAR DOCKING. SPECIFICALLY STIMULATES THE
CC INTRINSIC GTPASE ACTIVITY OF THE RAS-RELATED PROTEIN RAB1A AND
CC RAB5. SUGGESTING A POSSIBLE MECHANISM FOR ITS ROLE IN REGULATING
CC CELLULAR GROWTH (BY SIMILARITY).
CC -1- SUBUNIT: INTERACTS WITH HAMMARTIN. MAY ALSO INTERACT WITH THE
CC ADAPTER MOLECULE RABAPIN 5. THE FINAL COMPLEX CONTAINS TUBERIN
CC AND RABAPIN 5 LINKED TO RAB5 (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. AT STEADY STATE FOUND IN
CC ASSOCIATION WITH MEMBRANES (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 7 ISOFORMS: A, B, C, D, E, F AND G (SHOWN
CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED.
CC -1- SIMILARITY: CONTRAINS 1 RAB/RAN-GAP DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U37775; AAA6902.1; -.
DR EMBL: U37775; AAA6901.1; -.
DR EMBL: U39818; AAB18754.1; -.
DR EMBL: AB009371; BAA28845.1; -.
DR EMBL: AF132986; AAD27867.1; -.
DR MGD: MGI:102548; Tsc2.
DR InterPro: IPR000331; Rap_GAP.
DR Pfam: PF02145; Rap_GAP.1.
DR Anti-oncogene; Alternative splicing; GTPase activation.
DR Anti-oncogene; Alternative splicing; GTPase activation.
FT VARSPLIC 79 115 MISSING (IN ISOFORM A).
FT VARSPLIC 534 572 MISSING (IN ISOFORM B).

```

```

FT VARSPLIC 947 989 MISSING (IN ISOFORM F).
FT VARSPLIC 947 990 MISSING (IN ISOFORM C).
FT VARSPLIC 1245 1258 GHAPVIVATATGC -> RDTALYKSLSVPAAG (IN
FT VARSPLIC 1271 1293 ISOFORM E AND ISOFORM F).
FT VARSPLIC 1693 1728 MISSING (IN ISOFORM D, ISOFORM E AND
FT VARSPLIC 1693 1728 ISOFORM F).
FT VARSPLIC 1693 1728 GPACKCEWNRQPEIIVVVALPVNMLVTTLCHLO -> M
FT VARSPLIC 1693 1728 EGLVDTSVARIKSDRNLSFVAROMALHAN (IN ISOFORM
FT VARSPLIC 1693 1728 E AND ISOFORM F).
FT VARSPLIC 1693 1728 MISSING (IN ISOFORM F).
FT VARSPLIC 1693 1728 MISSING (IN REF. 2).
FT VARSPLIC 1693 1728 G -> S (IN REF. 2).
FT VARSPLIC 1693 1728 A -> S (IN REF. 2).
FT VARSPLIC 1693 1728 I -> N (IN REF. 2).
FT VARSPLIC 1693 1728 D -> N (IN REF. 2).
FT VARSPLIC 1693 1728 K -> N (IN REF. 4).
FT VARSPLIC 1693 1728 VP -> AA (IN REF. 2).
FT VARSPLIC 1693 1728 R -> P (IN REF. 2).
FT VARSPLIC 1693 1728 MISSING (IN REF. 2).
FT VARSPLIC 1693 1728 MISSING (IN REF. 2).
SQ SEQUENCE 1814 AA: 202069 MW: 913AB87194ADA5B CRC64:

```

```

Query Match 13.7%: Score 67.5; DB 1: Length 1814;
Best Local Similarity 27.5%; Pred. No. 41;
Matches 25; Conservative 17; Mismatches 36; Indels 13; Gaps 3;

```

```

QY 6 VGIOGSIHFHFKSKDELIRAVMEETI-----HYNTAMRASLEASTVREVRVLAIR 58
DB 513 VDLAEGCITHHFNLSLDIEKVMARSLSPELEERDLAVASLEEDVKTAVGLVILVQ 572
QY 59 CELQSIMGSGEAMAVLYEWMSLSAEGQAH 89
DB 573 TKLYTLPA---SHATRYVE---SLISHIOLH 597

```

```

RESULT 15
MTRR_NEIGO STANDARD: PRT: 210 AA.
AC P39897;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE REGULATORY PROTEIN MTRR.
GN MTRR.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19;
RX MEDLINE=94254732; PubMed=8196548;
RA Pan W., Spratt B.G.;
RT "Regulation of the permeability of the gonococcal cell envelope by
the mtr system.";
RL MOL. MICROBIOL. 11:769-775(1994).
CC -1- FUNCTION: PUTATIVE REPRESSOR OF MTRC GENE. CONTROLS THE
PERMEABILITY OF THE CELL ENVELOPE TO HYDROPHOBIC COMPOUNDS SUCH AS
ANTIBIOTICS AND DETERGENTS.
CC -1- SIMILARITY: BELONGS TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL
REGULATORS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```

DR EMBL: Z25797: CAA81047.1: -
DR PIR: S40250; S40250.
DR HSSP: P09164; 2RT.
DR InterPro: IPR001647; HTH_Tetr.

```

```

DR Pfam: PF00440; tetr. 1.
DR PROSITE: PS01081; HTH_TETR_FAMILY: 1.
KW Transcription regulation; DNA-binding; Repressor.
FT DNA_BIND 32 51 H-T-H MOTIF (POTENTIAL)
FT VARIANT 105 105 H -> Y (IN PERICILLIN-RESISTANT
FT ISOLATES).
SQ SEQUENCE 210 AA: 24192 MW: 4126446CBFE57F2 CRC64:

```

```

Query Match 13.6%: Score 67; DB 1: Length 210;
Best Local Similarity 16.6%; Pred. No. 3.8;
Matches 24; Conservative 25; Mismatches 32; Indels 64; Gaps 4;

```

```

QY 1 DLASAVGIOGSIHFHFKSKDELIRAVME-----
DB 35 EIAQAGVTRGALVWHRKNEKEDLDALFORICDDIENCIAODADAGSGVTVRRHLLH 94
QY 30 -----ETHY-----NTAMRASLEASTVREVRVLAIR--RCELQSIM 65
DB 95 FFERLQSDIHFKFHNILFLKCEHTEQMAAVIAIRKHOAIRREKITAVALTEAVENODLA 154
QY 66 GSGEAMAV-----LYEWMS 81
DB 155 DDLDEKETAVALFIKSTLDGLIWRMS 179

```

```

Search completed: March 4, 2002, 20:28:54
Job time: 97 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 4, 2002, 20:25:12 ; Search time 27.32 Seconds
(Without alignments)
535.404 Million cell updates/sec

Title: US-09-966-608-2

Perfect score: 493

Sequence: 1 DLASAVGIQSGSIFHHFKSK.....SLSAEGQAHVLAIRDVEEI 100

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_17:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	484	98.2	207	2	Q9RF4
2	429	87.0	198	2	Q9H2W2
3	110.5	22.4	190	2	O59306
4	110.5	22.4	197	2	Q9RRV9
5	94.5	19.2	200	2	P96839
6	92.5	18.8	194	2	P94548
7	92.5	18.8	215	2	Q9K8P5
8	90	18.3	192	2	Q9A8K8
9	90	18.3	222	2	O9S314
10	89	18.1	192	2	O67927
11	86	17.4	192	2	Q9K3Q4
12	86	17.4	195	2	Q9K8A4
13	85	17.2	197	2	Q9HTJ0
14	84.5	17.1	215	2	O9S253
15	84	17.0	206	2	Q9X7X0
16	83	16.8	214	2	Q9JN89
17	82.5	16.7	223	2	Q9X7X9
18	82	16.6	265	2	O9AA36
19	81	16.4	190	2	O32228

20	80	16.2	186	2	Q9I497	Q9I497	pseudomonas
21	80	16.2	212	2	Q9HZP1	Q9HZP1	pseudomonas
22	79	16.0	209	2	Q9XA31	Q9XA31	streptomyces
23	78	15.8	196	2	Q9RM14	Q9RM14	bacillus an
24	78	15.8	211	2	Q9A905	Q9A905	caulobacter
25	77.5	15.7	187	2	Q9K3M6	Q9K3M6	streptomyces
26	77.5	15.7	187	2	O53165	O53165	mycobacteri
27	77.5	15.7	194	2	Q9I1S1	Q9I1S1	pseudomonas
28	77	15.6	198	2	O66558	O66558	aquifex aeo
29	77	15.6	206	2	Q9A905	Q9A905	caulobacter
30	77	15.6	215	2	Q9A690	Q9A690	caulobacter
31	77	15.6	216	2	Q9K017	Q9K017	neisseria m
32	77	15.6	251	2	O9S261	O9S261	streptomyces
33	76.5	15.5	202	2	Q9A250	Q9A250	caulobacter
34	76.5	15.5	216	2	Q9JY29	Q9JY29	neisseria m
35	76.5	15.5	217	2	O87854	O87854	streptomyces
36	76	15.4	192	2	Q9Z6B7	Q9Z6B7	streptomyces
37	76	15.4	195	2	Q9X8M4	Q9X8M4	streptomyces
38	76	15.4	202	2	Q9KX78	Q9KX78	streptomyces
39	76	15.4	213	2	O9S250	O9S250	streptomyces
40	76	15.4	215	2	Q9AA74	Q9AA74	caulobacter
41	75.5	15.3	209	2	O9S2V6	O9S2V6	streptomyces
42	75	15.2	82	2	O07388	O07388	mycobacteri
43	75	15.2	190	2	Q9A109	Q9A109	staphylococ
44	75	15.2	193	2	Q9I209	Q9I209	pseudomonas
45	75	15.2	1489	5	Q9M1R3	Q9M1R3	drosofila

ALIGNMENTS

RESULT 1

Q9RF4 ID Q9RF4 PRELIMINARY: PRT: 207 AA.
AC Q9RF4;
DT 01-MAY-2000 (TREMUREL. 13, Created)
DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
DE PUTATIVE TRANSCRIPTIONAL REGULATOR PF12.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20349410; PubMed=10889151;
RA Sutton C.L., Kim J., Yamane A., Dalwadi H., Wei B., Landers C.,
RT Targan S.R., Braun J.;
RT "Identification of a novel bacterial sequence associated with Crohn's disease";
RL Gastroenterology 119:23-31(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Wei B., Huang T., Dalwadi H., Sutton C.L., Braun J.;
RT "12, a Crohn's disease-associated microbial gene and T-cell superantigen, is a species-specific product of Pseudomonas fluorescens";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO THE TETR/ACR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: AF173683; AAK16600.1;
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PRINTS: PR00455; HTH_Tetr.
DR PROSITE: PS01081; HTH_TETR_FAMILY.1.
KW DNA-binding; Transcription regulation.
SQ SEQUENCE 207 AA; 23107 MW; DD3AFB1FE10A736 CRC64;

Query Match 98.2%; Score 484; DB 2; Length 207;
Best Local Similarity 98.0%; Pred. No. 2.7e-42;
Matches 98; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 DLASAVGIQSGSIFHHFKSKDELLRAVMEETIHYNTAMRASLEFASVREVALIACE 60

```

Db 49 DLASAVGIQSGSIFHNRKSKDELIRAVMEETIHYNTAMMRASLEASTVREYVALIRCK 108
      |||
Qy 61 LOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDVEEQI 100
      |||
Db 109 LOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDVEEQI 148

RESULT 2
O9H2M2
ID O9H2M2 PRELIMINARY; PRT; 198 AA.
AC O9H2M2;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROBABLE TRANSCRIPTIONAL REGULATOR.
GN PA2885.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=2043737; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Salier M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
CB EMBL: AE004714; AAC06273.1; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PRINTS: PR00455; HTHTEtr.
DR PROSITE: PS01081; HTH_TETR_FAMILY; 1.
KM Complete proteome: DNA-binding; Transcription regulation.
SQ SEQUENCE 198 AA; 22057 MW; 079217CCCB8FF85 CRC64;

Query Match 87.0%; Score 429; DB 2; Length 198;
Best Local Similarity 84.0%; Pred. No. 1.2e-36;
Matches 84; Conservative 10; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DLASAVGIQSGSIFHNRKSKDELIRAVMEETIHYNTAMMRASLEASTVREYVALIRCE 60
      |||
Db 40 DLASAVGIQSGSIFHNRKSKDELIRSVMEETIHYNTALMRALADARDLREYGLIRCE 99

Qy 61 LOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDVEEQI 100
      |||
Db 100 LOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDVEEQI 139

RESULT 3
O59306
ID O59306 PRELIMINARY; PRT; 190 AA.
AC O59306;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S21.
OS Clostridium pasteurianum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1501;
RN [1]
RP SEQUENCE FROM N.A.
RA Meyer J.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
DR EMBL: Z28353; CAA82211.1; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PROSITE: PS01081; HTH_TETR_FAMILY; 1.
KM DNA-binding; Ribosomal protein; Transcription regulation.
SQ SEQUENCE 190 AA; 21692 MW; 0C6ED92A48AA69C0 CRC64;

Query Match 22.4%; Score 110.5; DB 2; Length 190;
Best Local Similarity 27.0%; Pred. No. 0.00085;
Matches 27; Conservative 25; Mismatches 35; Indels 13; Gaps 3;

Qy 1 DLASAVGIQSGSIFHNRKSKDELIRAVMEETIHYNTAMMRASLEAS---TVREYVAL 56
      |||
Db 28 ELASAGVAGTLYHNRKSKDELIRAVMEETIHYNTAMMRASLEAS---LKNNEIDATDKETLAKAV 83

Qy 57 IRCLEOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDV 96
      |||
Db 84 CRYOLNLIYKRNDFEKVIASQLMCK---ELROLELDI 118

RESULT 4
O9RRV9
ID O9RRV9 PRELIMINARY; PRT; 197 AA.
AC O9RRV9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TRANSCRIPTIONAL REGULATOR, TETR FAMILY.
GN DR2376.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1."
RL Science 286:1571-1577(1999).
CC -1- SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
CB EMBL: AE002068; AF11921.1; -
DR TIGR: DR2376; -
DR InterPro: IPR001647; HTH_Tetr.
DR Pfam: PF00440; tetr.1.
DR PRINTS: PR00455; HTHTEtr.
KM Complete proteome: DNA-binding; Transcription regulation.
SQ SEQUENCE 197 AA; 22307 MW; 769F2960C35F848D CRC64;

Query Match 22.4%; Score 110.5; DB 2; Length 197;
Best Local Similarity 27.5%; Pred. No. 0.00088;
Matches 28; Conservative 23; Mismatches 42; Indels 9; Gaps 2;

Qy 1 DLASAVGIQSGSIFHNRKSKDELIRAVMEETIHYNTAMMRASLEASTVREYVAL 56
      |||
Db 32 DLAGQGLMGQGLSLAHNSGKELLVEIVRGASQGFDEALFSLDYNLPADDEKLRANFRH 91

Qy 57 IRCLEOSIMGSGEAMAVLYEWMRSLSAEGOAHVALRDV 98
      |||
Db 92 IQ-----VVADNDSATVFPEHMKHLSAEPPYAQVAVARDTID 128

RESULT 5
P96839

```

ID P96839 PRELIMINARY: PRT: 200 AA.
 AC P96839:
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 22.9 KDA PROTEIN.
 GN RV357C OR MTCY06G11.04C.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
 RA Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 CC Nature 393:537-544(1998).
 RC -1 SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 DR EMBL: Z92774; CAB07159.1;
 DR InterPro: IPR001647; HTH_Tetr.
 DR Pfam: PF00440; tetr. 1.
 KW Complete proteome: DNA-binding: Hypothetical protein;
 KW Transcription regulation.
 SQ SEQUENCE 200 AA: 22906 MW: E1B29D39CF4D955F CRC64:

Query Match 19.2%; Score 94.5; DB 2; Length 200;
 Best Local Similarity 26.9%; Pred. NO. 0.04;
 Matches 21; Conservative 20; Mismatches 34; Indels 3; Gaps 1;

QY 1 DIASVVGSGSIFHHFKSKDEILRAVMEETIHYNTAMRASLEASTYREVALIRCE 60
 DB 35 DIADAGILSGSLYHNFASKEMVDLLRGFLDWLFARVDIVDSTANPLERLQGLFMA 94
 QY 61 QSIMGSGEAMAVLYVE 78
 DB 95 FEALIEHHAQ--VVIYQ 109

RESULT 6
 P94548 PRELIMINARY: PRT: 194 AA.
 AC P94548:
 DT 01-MAY-1997 (TREMBlrel. 03, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 22.0 KDA PROTEIN.
 GN YSLA.
 OS Bacillus subtilis.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;
 CC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=97217425; PubMed=9063446;
 RA Goethel S.F., Schmid R., Wipat A., Carter N.M., Emerson P.T.,
 RA Haywood C.R., Marahiel M.A.;
 RA "An internal FK506-binding domain is the catalytic core of the prolyl
 RT isomerase activity associated with the Bacillus subtilis trigger
 RL factor."
 RL Eur J Biochem 244:59-65(1997).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogatawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertoletti M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borriess R., Boursier L., Brans A., Braun M., Bignelli S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabre C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golligly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Kieffer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lamber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Malet C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogilwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porrolik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Toseuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wandut R., Wedler E., Wedler H., Weitzenger T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RA "The complete genome sequence of the gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Kunst F., Ogatawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/Genbank/DBJ databases.
 CC -1 SIMILARITY: TO THE TETR/ACRR FAMILY OF TRANSCRIPTIONAL REGULATORS.
 DR EMBL: Z75208; CAA99572.1;
 DR EMBL: Z95118; CAB14815.1;
 DR InterPro: IPR001647; HTH_Tetr.
 DR Pfam: PF00440; tetr. 1.
 DR PRINTS: PR00455; HTHTEPR.
 KW Complete proteome: DNA-binding: Hypothetical protein;
 KW Transcription regulation.
 SQ SEQUENCE 194 AA: 21978 MW: ABC01C9C3B47CFE8 CRC64:

Query Match 18.8%; Score 92.5; DB 2; Length 194;
 Best Local Similarity 27.4%; Pred. NO. 0.062;
 Matches 23; Conservative 22; Mismatches 38; Indels 1; Gaps 1;

QY 2 LASAVGSGSIFHHFKSKDEILRAVMEETIHYNTAMRASLEASTYREVALIRCEL 61
 DB 32 IAKQGVADGTYLVFKNNEDILISLFKMKGFIERMEDIKKATATKEK-LALVISKH 90
 QY 62 QSIMGSGEAMAVLYEMRSAP 85
 DB 91 FSLAGDHMLAIVTDLERQSNLE 114
 RESULT 7
 Q9K8P5 PRELIMINARY: PRT: 215 AA.
 AC Q9K8P5:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE TRANSCRIPTIONAL REGULATOR (TETR/ACRR FAMILY).
 GN BH2958.
 OS Bacillus halodurans.
 CC Bacteria; Firmicutes; Bacillus/Clostridium group;

DT 01-OCT-2000 (TREMBLRel. 15, Created)

DR PROSITE; PS01081; HTH_TETR_FAMILY; UNKNOWN_1.

KW Complete proteome; DNA-binding; Transcription regulation.

Tue Mar 5 07:57:19 2002

us-09-966-608-2.rspt

Page 7

